March 2006

Measuring and Understanding Connectivity

Brendan L. Wilson
Worcester Polytechnic Institute

Liam Macquarrie Morley
Worcester Polytechnic Institute

Matthew Kieren Brennan
Worcester Polytechnic Institute

MEASURING AND UNDERSTANDING CONNECTIVITY

An Interactive Qualifying Project Report:

submitted to the faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Bachelor of Science

by

Matthew K. Brennan

Liam Morley

Brendan Wilson

Date: March 02, 2006

Approved:

Professor Craig Wills, Project Advisor

1. connectivity
2. transportation
3. Internet

Abstract

This report examines the correlation of Internet travel time to real-world travel times, and also compares the growth of the Internet to the growth of other technologies that connect people. In compiling a series of comprehensive datasets, and utilizing scripts to gather specific Internet and physical data, this project examines the impact of trends in connectivity on a local and national scale. Also investigated are the effects of routing on both forms of connectivity, as well as other influencing factors.

Acknowledgements

We would like to sincerely thank those who have assisted us during the course of this project. In particular, we thank our project advisor, Professor Craig Wills, for his help and guidance along the way. During the course of the project, we made extensive use of CAIDA, the Google Maps service and API, MapQuest, and Expedia.com. Without these services (and in some cases, their associated communities), we could not have conducted our project.

Table of Contents

List of Figures

List of Tables

1 Introduction

Within four decades, the Internet has grown from a simple network involving two computers to a series of worldwide interconnections linking not just computers, but people, businesses and cultures. This rapid growth has prompted many to look at and investigate different aspects of the systems involved, and how each area of the Web is connected. Our project is focused on Internet travel time, or the time it takes data to reach and return from its destination, be it across the hall, or across the country. This report looks to examine the correlation of this time to real-world travel times and distances, and also compares the growth of the Internet to the growth of other technologies that connect people.

By collecting and examining the data in these areas, we have determined that there exists a correlation between the differing types of transport. In this report, we also explain any data points that deviate from the expected. We also explain correlations between non-trend datasets, including all geographical and travel information we have collected. We found maps of U.S. Internet backbones through our research. We then took the data from these maps to better understand the backbones, and then compared these backbones to commonly traveled routes for both automobile and airplane traffic systems. This, we hope, will allow us to gain a greater understanding for how the Internet distributes data across the country and around the world, and also allow us to understand how the Internet compares to other forms of transfer involving both people and data across hundreds of miles.

In order to best understand the relationship between modern day real-world transportation systems and the Internet, it is important to understand the similarities and differences in the ways that these networks evolved over time. What made the system necessary? What came before it? How fast did it grow, and how quickly was it absorbed into society? Understanding the answers to these questions aid us in understanding how these systems are related today.

A comparison between travel times in physical-space versus cyberspace was the original concept behind our IQP. Professor Wills presented some simple graphs and possible correlations between physical-space and cyberspace with the attached phrase “you could do a whole lot with this”. Our collective curiosity was piqued and we became interested with what similarities the two distance measurements shared, and where they differed. . By looking at the routing data for the Internet, developing metrics of measurement for that data, and examining the outliers of our data, we have come to a better understanding of just what similarities really do exist between our two focus areas. Through this project we examine the degree of correlation between physical-space and cyberspace and to go even further than the project was initially intended. By looking into different areas of both the Internet and physical space, we hoped to refine the project to gather the data we were looking for.

Routing is a huge concern in the world of data packet travel. Finding the quickest way from node A to node B often entails algorithms that use information on possible routes. Physical-space travel paths are a little different, largely because we move much slower and are more deeply impacted by the geography of our environment. The only times we will choose to avoid a highway or road is when it is completely backed up. The reason for this difference is due to the impact of queuing on each traveler. A packet can remain queued for a portion of time equivalent to the time it takes to actually travel from one node to the next. In a car, it would be a very rare occurrence to be deadlocked for longer than the journey would have taken. To help us with measuring the Internet data that we collect, we will have to implement some type of metric which will define what data we are looking at and how it compares across the different routes.

There are a large number of possible metrics that we considered using to measure the data we gathered on Internet connectivity. From these metrics, we narrowed the field down to one or two heuristics (discussed in Section 3.3) which clearly state the information we are looking for, and are better able to compare our Internet data with the real world data we collect. CAIDA (the Cooperative Association for Internet Data Analysis) is a group that has already done a great deal of research in this area, and has many detailed papers describing what metrics are the "best" or provide the most accurate results.

As we progressed with our connectivity data collection we have addressed to some degree the problem of different types of connectivity in different areas. While WPI has a high-speed connection with the Internet, a normal person using a commercial Internet Service Provider (ISP) could have a much slower connection comparatively. This difference could be caused by a variety of factors, from a different type of connection, to distance from the major backbones.

Although finding correlations between physical space and cyberspace measurements of time and space is a strong focus for this IQP, there might actually be more intriguing characteristics in the outliers of the data we collect. If we find that 95% of the data we collect follows along a linear trend, then what about the other 5%? What causes a specific node to have good Internet connectivity, but poor travel connectivity?

Will there be outliers that we can drive to in a matter of minutes, but have round trip times as if they were in California?

To aid us in our search for data to collect, we began a search for research that has been performed before on Internet connectivity. Specifically, ready-made programs and scripts were searched for which would potentially make our job easier. We were again pulled into the CAIDA website, as they have a comprehensive list of different tools that they and other educational institutions use to measure Internet data. From globally mapping all the backbones to providing route maps and Round Trip Times (RTT), these tools cover most of the data we gathered on Internet connectivity. Their tools stretch from highly advanced, like the Skitter mapping program, to a simpler program, TraceRoute, which simply finds a route from one address on the Internet to another.

Using the ideas discussed in our here, we began to reach out and explore the different avenues of research that have seemed to be of interest to the project. As will be explained in the following chapters, we examine each of the topics in-depth, and explain what our focus for the project in each area of study. Using the methods recorded below, we began to gather and analyze data and draw conclusions from it.

Our report will adhere to the following outline. First we will begin our report with a discussion of the motivations behind this project, why we are doing it, and what inspired us to work on it. Second, we delve into the background of our topic and explain the work that we found has been completed up to this point. Next we will talk about the datasets we collected, and the procedure we used to gather this data. With the completed datasets, we will present our results and analyze them. Following this, of our report will include our conclusions about our analyzed data. We hope to draw conclusions in regard to data trends, social implications of connectivity, comparative connectivity, and the effects of routing. Finally, we will offer different paths in which further work could be completed in this area of study.

2 Motivation

Within this project, we relate cyberspace measures of connectivity with real-world measurements of connectivity. We find not only interesting data in doing such a comparison, but also make significant observations from our work. Our interest in the topic of relating cyberspace to physical space has resulted in us exploring several different paths of research.

When we first looked into this project, we looked at simple maps that compared 'connectivity'. We began with the desire to focus simply on gathering data and measurements with the goal of comparing round trip time to real world travel time. As we conversed more with Professor Wills, we kept coming up with new fascinating questions that we wanted to answer. Part of the task of defining the project was narrowing its scope to answer only the questions pertaining to our primary interest. We have listed a few additional questions worth answering in Chapter 7, our future work chapter.

One of the largest areas of interest regarded data that might not fit into any trend that we discover. We expected to find that, for the most part, our data would fit to a linear regression in which one axis would be physical-space connectivity (measured in units of travel time) and the other axis would be RTT. We also posited that there would be some points that did not match up well to a linear regression and that these particular points might provide for even further insight into the nature of connectivity.

While thinking about the comparisons we were considering, we wondered what other comparisons we could do. One point of curiosity was to compare RTT to driving time. We were interested in seeing if RTT would correlate more strongly to driving time or cumulative travel time (which is the sum of driving time and air travel time). We were also interested in how Internet backbones affect Internet connectivity, as well as how air travel affects real world connectivity. Another point of curiosity involved how routing behavior and routing characteristics might impact connectivity. With all of these questions lingering, we began looking into the evolution of various types of connectivity including planes and the Internet. We also looked into comparing real world travel paths with data packet routes.

Finally, with all these propositions in front of us, we were driven to find what tools were available to us for use in gathering data. Just as a carpenter is only as good as his tools, we would need to find software that would not only allow us to efficiently retrieve data, but also give us data on which we could easily perform calculations.

To a certain extent, we have answered questions that some have asked before. Our motivation for this project has been to ask new questions that no one has answered, such as certain trends that we can glean from analyzing a large set of data from a geographical standpoint. We hoped to expand our own knowledge and understanding both in the subject of connectivity comparison as well as connectivity in general, and we feel that we have succeeded in this area. We also hope that our project can offer food-for-thought to others interested in similar research matters.

3 Background

We began this project by researching different aspects of our connectivity and evolution questions. This section forms the basis of those investigations, and the knowledge we eventually used to investigate an area of Internet behavior and patterns that only a few people have begun to explore.

3.1 Historical Change

It is important to understand the historical impact of changes in technology on transportation and connectivity. How have changes in American transportation, as well as changes to technology, affected the way that society is interconnected?

3.1.1 Transportation History

In order to understand these changes, it is important to understand how American transportation has changed over time. In *The American Transportation Problem* [14], Harold Moulton charts in depth the development of transportation from the birth of the Erie Canal, through the development of the railroad system, to the infancy of the airlines. Moulton points out that many railroad routes were built using canals that had been filled in. There is a clear relationship between one mode of transportation and its successor.

William R. Black offers additional background information regarding the growth and development of American transportation in *Transportation: A Geographical Analysis*. Black covers water transport, railroad development, the highway system, urban transit, and air transport. He makes an interesting comparison of the roads built in America to those built by the Persians, Greeks, and Romans; roads built by the latter three are more likely to be straight than roads built by the former. These patterns are due to the fact that many roads built in Persia, Greece and Rome were designed to be as straight as possible, while early land routes in North America were based off of Native American roads. These roads were designed using a “principle of least effort”, which dictates that roads should be built around existing paths whenever possible. These paths typically are built to avoid obstacles. Such information could also be used to explain connectivity differences between a colonial city like Boston and a planned, grid-like city like New York City. [2]

Black also correlates route development with trade flow and geography. Finally, Black discusses several societal trends and their duplex relationship with transportation. He states that, because of improvements to transportation technology (notably, the automobile), it is now much easier to travel distances in a shorter time. These improvements have contributed to urban sprawl. Black also discusses the impact of transportation technology on those living in rural areas. During the 1800s, an injury on a farm could result in death due to the delay in transporting the injured to a hospital. This is of course no longer the case. [2]

Wikipedia contains some helpful information on transport, including a relationship between the population density (the number of people needed to be transported) and the necessary mode of transportation, called Floor Area Ratio (or FAR). This is a relationship between the space that the transported objects take up and the space required by the medium to transport them. FARs of 1.5 or less are generally suited to cars, and FARs of six or more are generally suited to trains [18]. As the population of a city grows over time, it becomes necessary to improve the city's transportation infrastructure.

3.1.2 Telephone History

Frances Cairncross writes in her book *The Death of Distance* about the effects of communications media on society and industry. In her book, she chronicles the effects of the telephone on society. In 1927, a three-minute phone call placed from New York to London cost approximately $250 (in 1990 U.S. dollars, accounting for inflation). Such a call today would cost pennies. Improvements in technology, which allow the network to be built and operated at a reduced cost, are the reason for the large difference. Privatization of the industry also contributes to lower costs. With this reduction in cost over time comes a larger network; the less it costs to build a network, the further it can expand. [3]

This cost reduction has taken a novelty medium and transformed telephone networks into a communication standard. As of 1997, the telephone was three times more likely to exist in the home than a personal computer, even in developed nations [3]. The telephone eventually became a foundation for Internet connectivity, and as such helped to define the way that the Internet would grow.

The growth of networks makes it possible for poorer countries to catch up, and allows frequent communication across the globe. Asia is adding new telephone lines at a rate of 20 to 25 million per year; this number is roughly three times the number added per year in the United States [3]. Network growth also enables those in poorer countries to access the Internet, as a phone connection to the Internet is much more likely than a broadband or DSL connection.

Another improvement in telephone technology with major effect is that of wireless telecommunication. Cell phones have quickly infiltrated society, making up one in seven of the world’s telephone subscriptions in little more than 10 years [3]. Cairncross speculates that cell phones will be as ubiquitous as a wallet or watch, and that they may become built into hybrid items [3]. The latter has already come to pass, with such inventions as blackberry personal digital assistants and camera cell phones.

3.1.3 Internet History

In researching Internet growth, one helpful resource was an online exhibit on the Computer History Museum website, “Internet History”. The exhibit charts Internet growth from its early conception in 1962 to when the World Wide Web is in high gear, in 1992. The exhibit briefly discusses several key moments in time, such as the first network, constructed in September of 1969 by the Advanced Research Projects Agency (ARPA). This network was first referred to as ARPANET. The exhibit also includes growth charts, outlining the introduction and maturation of backbones. [5]

Several aspects of Internet evolution, as viewed in a cartographic sense, can be found in Martin Dodge and Rob Kitchin’s *Atlas of Cyberspace*. Dodge and Kitchin include a large number of maps which chart several areas of Internet growth, including backbone connectivity, worldwide growth of Internet services, and social networks established purely on the Internet. Such networks include e-mail, Usenet, and even virtual worlds such as EverQuest. One map from 1997, which is largely geographical, points out that although most of the world has acquired Internet access, there are still a few remote third world countries that have not. Another such map shows the distribution of domain names for the San Francisco area and Silicon Valley, which clearly is related to urban development. Many of the maps in the book have references to further information on the Internet. [6]

Cairncross' *The Death of Distance* is also beneficial for explaining the effect of the Internet on society. Cairncross cites e-mail, the ability to search for information efficiently, and other such services as some of the more useful products of the Internet. Though the U.S. dominates Internet usage, accounting for more than 50% of its users, the areas of the world that are enjoying the fastest development are elsewhere. Cairncross speculates that if the Internet can reach global penetration, it may become the primary method for worldwide communication and supersede or envelope other forms of media. [3]

3.1.4 Relating Changes in Transportation to Changes in Connectivity

As time has advanced, improvements in transportation technology have made it possible for people to communicate faster with others. Likewise, improvements in telephone technology have increased one's ability to contact others more frequently, as well as broadened one's range of potential contacts. The Internet has also resulted in an increase in personal interaction. With the Internet, people can transfer information almost instantly. In addition, due to such technologies as e-mail and "instant messenger", people can contact one another more frequently due to the reduction in cost of communication. As the Internet grows in size, the social network will grow with it. Understanding these changes over time and how these changes relate to each other will give us more depth of understanding during the course of our research.

3.2 Distance to Round Trip Time Correlation

The method for determining whether or not a computer can be connected to from a network is called “ping”. If a person at computer A wanted to see if computer B were available through a network, that person would send a signal from computer A to computer B and wait for a response. If a response came back, then computer B would be able to be “pinged” from computer A. Otherwise, it would be unavailable. The time between sending a ping and receiving a response is round trip time.

There is an undeniable correlation between distance and RTT; this has been established repeatedly in such papers as G’unther and Hoene’s *Measuring Round Trip Times to Determine the Distance between WLAN Nodes* [9], in which they attempted to relate ping time to distance in a wireless network using consumer wireless cards. In Lepak and Crescimanno’s article *Speed of Light Measurement Using Ping* [12], the authors investigated using round trip travel time of data packets to interpolate the speed of an electrical signal in a transmission line along the network. As this was to become the key metric of our investigations, we needed to make sure that we understood exactly what it meant and what had been investigated before.

3.3 CAIDA Metrics

An important aspect of measuring connectivity is to understand the relationships between data travel time and real world travel time. Is there more than simply a correlation or are there trends and patterns? How should we measure these relationships? CAIDA has released a paper that describes four different metrics that would help us analyze and track this data.

The first, Internet Protocol (IP) path length, is basically a count of the number of hops a packet must take to reach its destination, which can greatly influence the round trip time for a bit of data. The second is Autonomous System (abbr. AS) path length, which uses single or multiple networks, and tracks the total number of ASes that a packet visits. The main benefit of this metric is that it works at no extra cost or strain on the network. Third, the geographical distance metric, looks at what the actual physical distance is via latitude and longitude. This metric compares the straight line distance to from target to the actual travel distance the packet takes as it is routed through the backbones and other networks to reach its destination. The fourth and final metric, is Round Trip Time, or RTT. RTT is a measure of several different metrics. The metric takes either the last RTT recorded or the median RTT for a set of values previously recorded. This metric is one of the most closely related to what we are able to visualize, and thus is widely used. However, it is also prone to many small factors which can greatly affect the result, including link congestion, queuing, and routing changes. [11]

3.4 Measurement Tools

Much of the information that we must investigate has already been collected, and there are dozens of programs on the CAIDA website that track different types of potentially useful data. The organization offers four categories of tools: measurement, utilities, taxonomy and visualization.

Some of the following programs in the measurement category could potentially be useful. Skitter, one of CAIDA's most heavily used programs, is a system that provides a wider view of the Internet. It measures IP paths, RTT, tracks routing changes, and is capable of visualizing the network connectivity on a global scale. [4]

Another potentially useful program is Beluga, which provides beginning-to-end RTT as well as the data for each of the hops the packet makes to get to a destination. This program also provides detailed and colorful graphs, which could have been useful when we compared data trends.

A third and final possibly useful program is GTrace, which is a graphical front-end user interface for TraceRoute. TraceRoute is a program that tries to determine each node a packet travels through as it heads towards its final destination. GTrace uses a series of decision-making algorithms to find and plot the locations of each node in the path of the data being analyzed [16]. This program could be useful for tracking Internet backbones or comparing routes to physical paths traveled over land.

3.5 Routing

Another point of interest is the difference in packet routing versus the travel path of human transportation. One article, *An Adaptive Routing Algorithm for In-Vehicle Route Guidance Systems with Real-Time Information* [8], discusses the most efficient and effective method for determining an optimal route for a given vehicle through a traffic network. The travel time on each link can be modeled as a random variable and its realization can be estimated in advance and made available to the vehicle's routing system before it enters the link. Another article, *Routing* [17], generally outlines some of the more common aspects of routing data across computer networks. As Wikipedia states, Routing is the means of discovering paths in computer networks along which information (split up into packets) can be sent.

3.6 Last Mile Connectivity

There are many factors that influence a computer's connection to the Internet. One of the major problems that is faced by a user is the "Last mile connectivity" of his or her service provider. Internet service providers are constantly faced with finding a way to set up their networks in the most cost-efficient and high-speed manner. The last mile connectivity is the hardest area for companies to connect, as it involves running numerous connections to computers off of the major lines. This process is expensive, and causes major headaches amongst the ISP community. [7]

The evolution of this type of connectivity has caused speed of Internet connections to increase dramatically. Starting with standard copper telephone wires, the systems have evolved into cable Internet, T1 connections, and the next generation of connectivity, dark fiber. The success of these different connections is dependent upon the ratio of cost compared to the speed at which they can connect with the major backbones of the network. All of this change has occurred as Internet Service Providers began seeking to increase speeds for their customers at cheaper rates. [7]

4 Methodology

4.1 Datasets

When collecting our data, it was important that we found data points that were accurate. If our data were to include an Internet location that was not located where we believed it to be, then we could have no confidence in our results. To aid in providing accurate data, we set out to narrow the range of the data we would collect to servers we had a high confidence in.

One of the constraints we imposed on our data set was to only select servers that belonged to universities. Universities typically have only a single location, which reduces guess work; state universities typically have multiple locations, and can be eliminated. We theorized that a server for a university would have a higher probability of existing at the university than being hosted elsewhere. This situation contrasts with servers for companies, which are often hosted by hosting companies that can be located anywhere. We also found it easier to compile a list of university data than company data.

We also chose to limit our data strictly to points that exist within the continental United States. One of the reasons for this is the fact that servers in the United States are more often online and available [1]. In addition to reliability, we also could more easily visualize the relation of Internet connectivity to physical connectivity on a national scale. It would also be easier to collect data on American airports, roads, and universities. When choosing the information to collect on each university, we created a metric called cumulative travel time (CTT). This is the sum time it takes to drive to an airport, fly to the airport closest to our destination, and then drive to our final destination. Cumulative travel time might also include connecting through intermediate airports. The CTT is in most cases the quickest route. There are times when a given location is so close that it makes more sense to drive there directly than it does to go by plane. In these cases, we used driving time in place of cumulative travel time. We chose this measure of travel due to the fact that the Internet also chooses the fastest path in all cases. In most scenarios involving long distances, information traveling across the Internet will use faster more direct pipes, called backbones. These backbones are similar to airline flights: they cover long distances (compared to local traffic routes), usually with a direct path, with few exits in between end-points.

There were a number of other important attributes that would chose to describe each university. These included the web address of the server; the zip code it was located at; its latitude and longitude; its nearest airport; the time taken to drive from WPI to the university; the cumulative travel time to the university; and its raw round trip time. We also inserted another attribute that represented the minimum round trip time based on several connection attempts. We preformed connection attempts from both on-campus and off-campus, and split the results into separate datasets. Through the use of scripts described in Section 4.2, we then gathered the data we required to draw conclusions about our topic.

In order to organize our data, we planned to divide up the raw collected data into four datasets; two for on-campus and two for off-campus. Two of the datasets, one on-campus and one off, would cover a nationwide sampling of data. The other two sets would cover a more focused sampling in just the New England area. For each of the New England datasets, we chose additional points in states in the New England area and eliminated all other points outside of that area. We further investigated the New England area to view our data on a more microscopic level. We also were interested in comparing various regional patterns of the area to the larger, national patterns.

The respective names of each dataset are: the off-campus nationwide dataset, the on-campus nationwide dataset, the off-campus New England dataset, and the on-campus New England dataset. By dividing our data into these sets, we hoped to be able to compare the differences between on-campus and off-campus connectivity.

4.2 Scripts

The basis of the data gathering methods used were online programs, available to the general public. The first program that we used for the project is MapQuest. Of the three potential programs we looked at, GoogleMaps, RandMcNally.com, and MapQuest, MapQuest allowed us to easily create scripts because of the verbose URLs used to communicate with the MapQuest servers. The ease of URL modification greatly aided our geographical data collection, and reduced the amount of additional overhead.

Using scripts similar to the one that interfaced with MapQuest, we also collected data on the travel times for airplanes. We created a set of the major commercial airports in the United States, including at least one from each state. We focused our efforts on the major airports and looked at the connections available between the different airfields listed on the FAA website. Using Expedia.com, we generated the minimum travel times from each airport to Logan Airport in Boston. Logan Airport was determined to be the most accessible airport for Worcester, and therefore for WPI.

We implemented a script which allowed us to enter a university we wished to get travel data for and the script returned the total travel time from WPI to the target university. The script first found the driving distance to the closest airport in our pre-defined set of airports. Then, it was just a matter of summing the travel time from WPI to Logan Airport (which is constant), to the airport closest to the target university, and the travel time from that airport to the target university. Using this method, we were able to quickly and efficiently gather the desired information on travel time.

4.3 Traffic

When considering travel, traffic plays an important role. This fact holds true in both real world travel and cyberspace travel. In the case of the real world, traffic can effect how long you have to wait at a light, how congested a road is, the wait time at an airport, or even seat availability on a given airplane. In the case of cyberspace, traffic can affect how long data is queued at any given node. Traffic can also notably effect the overall travel time of data. If a node chosen by our routing algorithm is forced to handle data from a video stream or a bandwidth intensive video game, our data packets will have to wade through these extreme waves of data.

Due to the variability in traffic and the difficulty in developing methods to compensate for traffic, we ultimately decided to disregard the effects of traffic on our data. In order to do so, we had to make a few assertions. First, we asserted that the data returned form MapQuest and Expedia did not factor in current or average traffic conditions for routes. Second, we asserted that the effects of traffic on RTT could be lowered to a negligible amount by performing several connection attempts and taking the minimum.

4.4 Summary

Through the development of our scripts, and the selection of specific metrics of measurement, we were able to gather large numbers of interesting data. Many of the universities gave us surprising results, and caused us to modify and change our datasets to replace problematic data. If the university seemed to not be located at the same place as its website, we removed it to provide us with more accurate results. By disregarding traffic, we hoped to remove some of the potential variability in our datasets, and give us a strong base from which to start our analysis.

5 Results and Analysis

Using the mentioned scripts, we gathered data to plot the different routes of airlines and cars along with Internet data on one map. Having such data allowed us to compare and contrast the differences between the different types of connectivity for each of our datasets. Viewing such a comparison as a map graphically showed the connectivity of each type of transport. These maps also showed what geographical areas our data covered. This type of visual comparison and analysis led into our development of our trends and a comparison of the non-trend data. Such a comparison also allowed us to find a reason as to why the non-trend data exists.

5.1 Trends

Before we began gathering data for our study, we had certain expectations. There were certain trends that we assumed we would see in our data. Our major assumption was that the further a destination away is, the longer it should take to get there, as this is the nature of travel. We considered this logic to be common sense, and anticipated seeing it expressed in our datasets. The focus of this project, however, was to look at the effects of connectivity. How well would these trends be conveyed?

The first data set to look at is the on-campus-nationwide dataset (Figure 5.1). In this scatter-plot, the x-axis represents cumulative travel time, and the y-axis represents RTT. Included for visual analysis is a linear regression of best fit. Looking at this graph, we vaguely see what was expected. Most of the points loosely fall on our trend line.

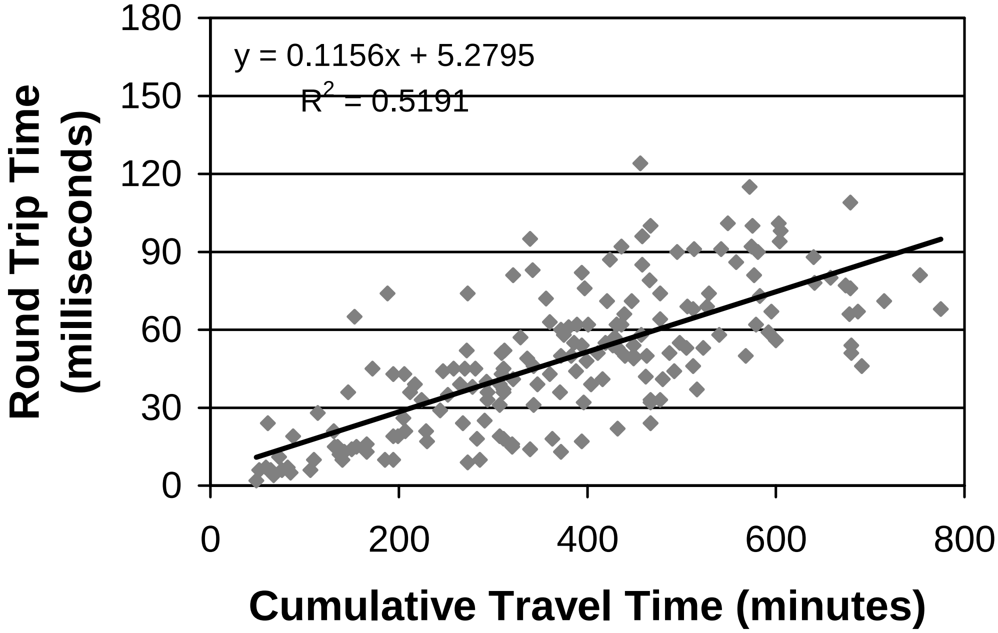

Figure 5.1 National Dataset, On-Campus Scatterplot

Is “loosely” good enough for us to establish this as an acceptable trend? It is important to consider the variable nature of the data we are discussing. Connectivity is dependant on many variables; traffic being one of the most important. When comparing two metrics that have inherent variability in them, a perfect fit to a trend line would be highly improbable. Even though our scatter plot’s correlation is lower than we would hope for, it is perfectly acceptable due to the nature of the data. We will discuss the statistical points of interest in Section 5.3.

Next we will look at the same data set pinged from off-campus (Figure 5.2). Our linear regression is now more centered because the extreme data points were truncated due to time-outs. Once again, we see a moderately strong linear correlation where locations further away seem slightly longer to respond. An additional trend visible in the last two scatter-plots can be seen. As a target is further positive on the x-axis, there is more variability on the y-axis. Any of the points that are roughly an hour away will only vary about 20 milliseconds on the y-axis. Meanwhile, points that take 10 hours to travel to will vary well over 150 milliseconds. This trend can be a result of additional factors such as traffic or can very well be expressing the connectivity variances we are focused on. This trend will be discussed further in section 6.

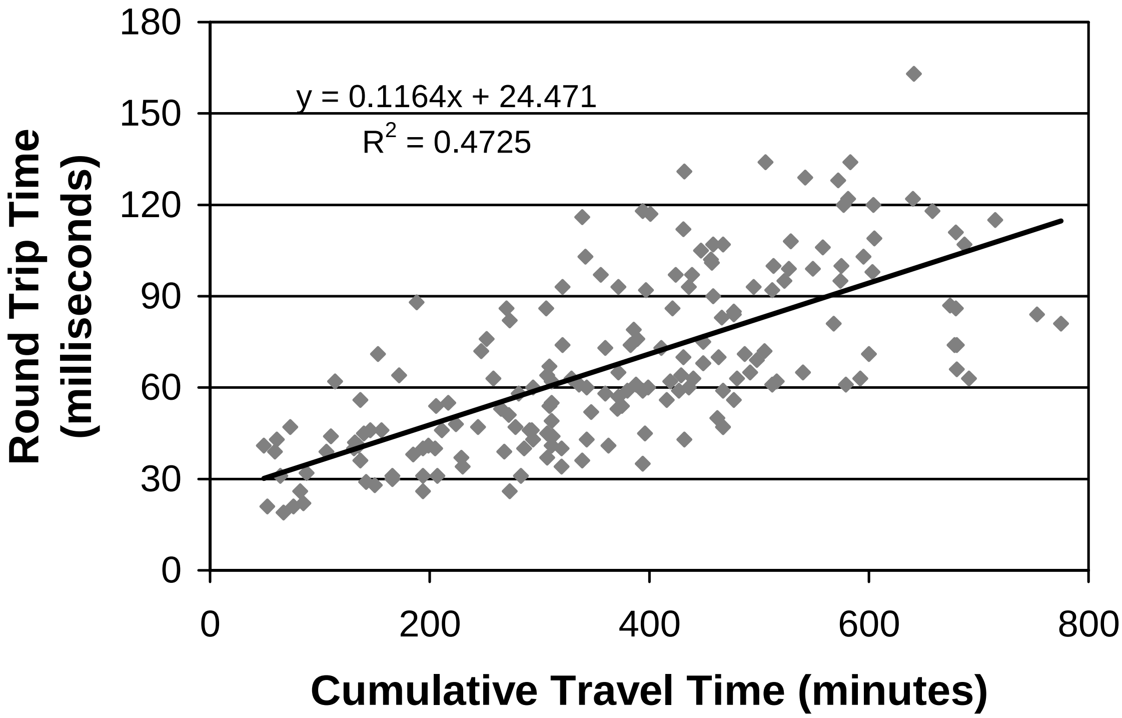

Figure 5.2 National Dataset, Off-Campus Scatterplot

Now we have two localized datasets to look at. Our New England data set was also pinged from on-campus (Figure 5.3) and off-campus (Figure 5.4). These points cover much less of a range on the x-axis, and therefore give us a more zoomed-in look at our data. Neither of these datasets adheres to any noticeable linear trend. We are unsure of the exact reason why we did not see any linear trends on these datasets. An additional point of interest is on the off-campus set. We see obvious ping time quanta, where any one of the RTTs are one of six different levels. This phenomenon might be attributable to a networking anomaly of which we are unaware.

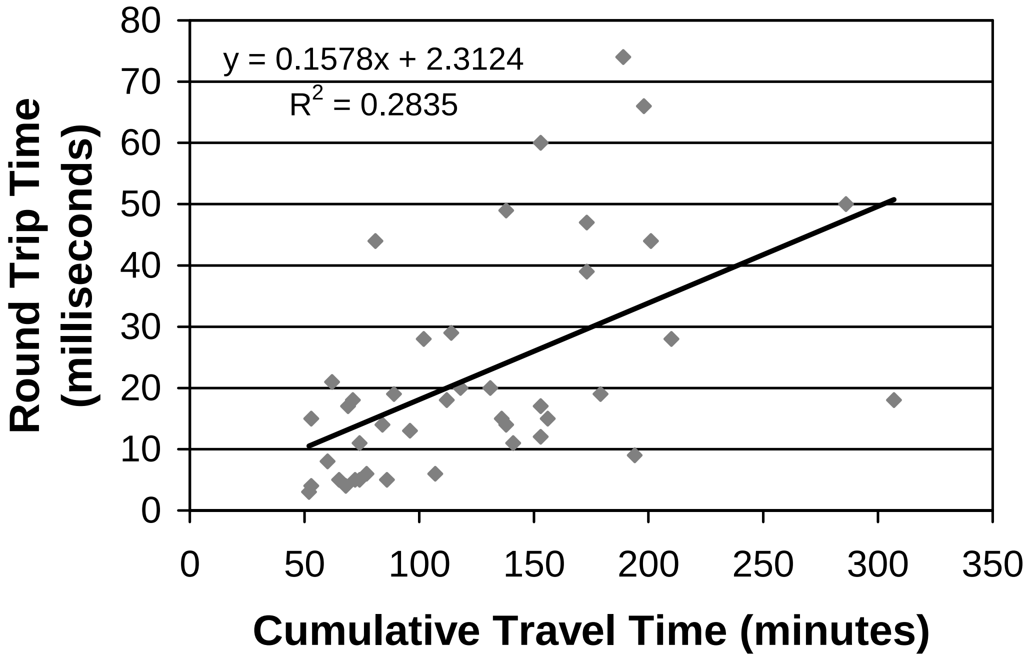

Figure 5.3 New England Dataset, On-Campus Scatterplot

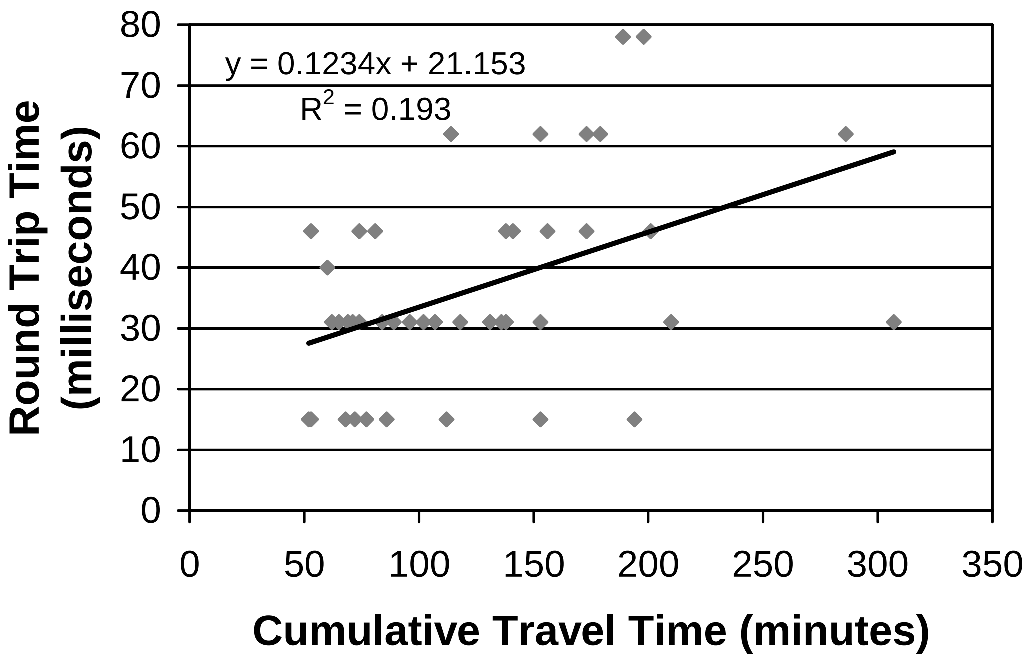

Figure 5.4 New England Dataset, Off-Campus Scatterplot

5.2 Non-Trend Data

Although much of our data falls roughly along a direct correlation between RTT and CTT, many points are not strongly correlated to the trend line. Non-trend data occurs whenever the relationship between physical travel time and RTT for a given location is out of sync with the same relationship for other locations.

As can be seen in the national on-campus dataset (Figure 5.1), several points clearly fall far away from the line of correlation. Most of these points are cases where it takes much longer to reach them on the Internet than it does physically. There can be many reasons for this situation. It could be that a given location is far away from any Internet hub or backbone, or it could also be that the point is abnormally well connected as far as physical travel is concerned. This behavior can be analyzed further to some degree by looking at additional factors. If the distance or driving time is significantly less than other points with a similar cumulative travel time, this could suggest that a person at this point can travel easily to an airport. Otherwise, the point can be assumed to be poorly connected to the Internet. Specific points will be discussed later in Section 5.4.

In comparing non-trend data between the on-campus and off-campus nationwide datasets, it is somewhat clear that there is less non-trend data in the off-campus nationwide dataset. The non-trend data points in the off-campus nationwide dataset are far from the line of correlation in the on-campus data set as well. In general, the large dispersion in the off-campus data set limits the number of non-trend data points. The geographic locations of these points can be viewed by plotting these points on a map.

Figure 5.5 On-Campus National RTT Data Superimposed on Map of US

The map in Figure 5.5 depicts different colored dots based on the round trip time for each location. Such a map gives a general impression of Internet connectivity.

As can be seen in Figure 5.6, there are many similarities between the on-campus results and the off-campus results, but a few differences. Each point responded more slowly when connected to from off-campus than on-campus. There seems to be a faster connection to the west coast in the off-campus results than in the on-campus results, whereas the on-campus results produce faster round trip times for the New Mexico region.

Figure 5.6 Off-Campus National RTT Data Superimposed on Map of US

When the statistical analysis from scatterplots is also shown on the map, points that fall far from the line of correlation can be easily spotted. White points designate locations that have a much smaller round trip time relative to their physical travel time, while black points have a much higher than the expected round trip time. Any point within a certain range of acceptability (a certain number of milliseconds faster or slower than the expected round trip time) is not displayed. In Figure 5.7, all points that are within 30 milliseconds of their expected RTT are not displayed. Such points fit closely enough to our trend line that they are not of interest in this map.

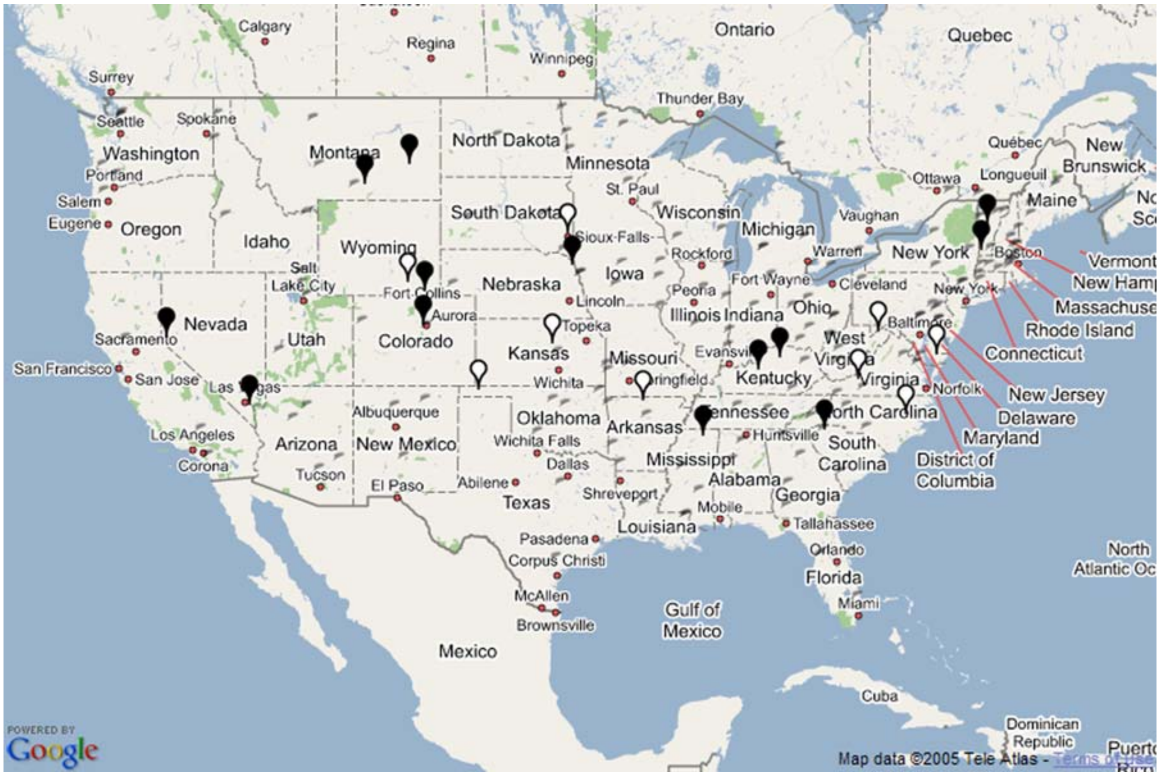

Figure 5.7 On-Campus National Dataset with Geostatistical Analysis (30 ms Acceptability Range)

As discussed earlier, most of the points in the New England datasets, both on-campus and off-campus, do not seem to follow any strong trend. This behavior is depicted geographically in Figure 5.8. There could be many different reasons for this. It could be simply a matter of resolution: perhaps because the range in ping times is so small, the differences only appear larger than they are. It could also be due to the fact that points in New England are so close together that distance does not have much chance to play a factor.

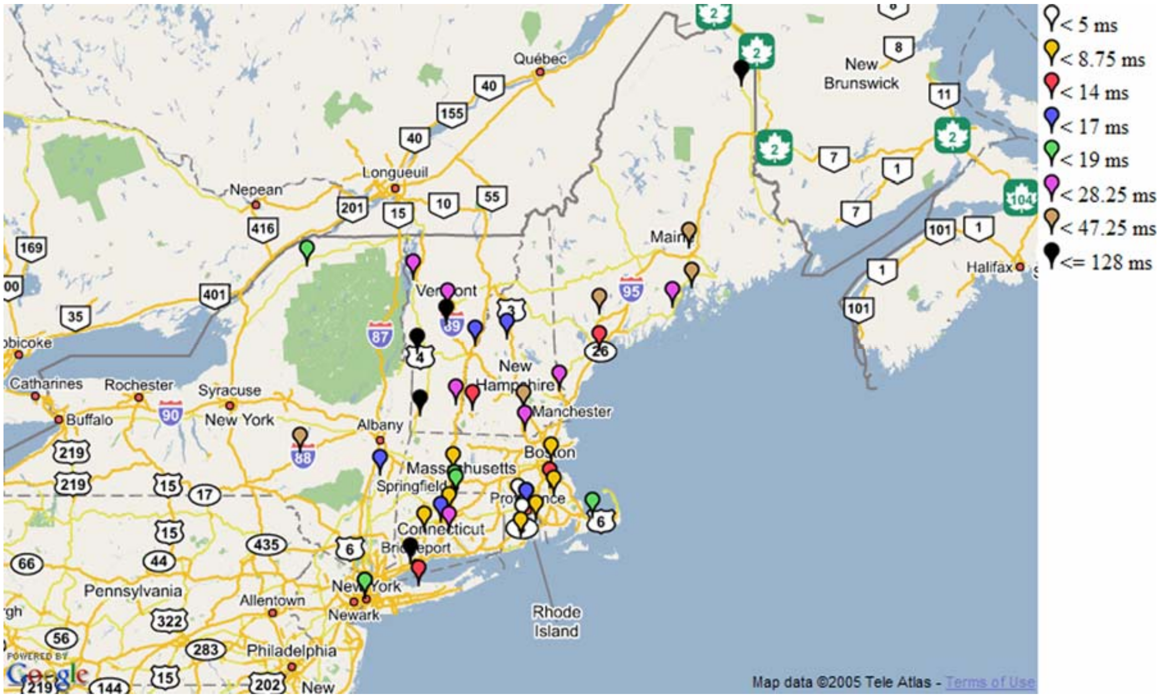

Figure 5.8 New England On-Campus Dataset (Geographic)

When looking at a geographical plot of the New England data, black dots (indicating a slow round trip time) can be found largely in Vermont and northern Maine, although Burlington Vermont is still well-connected. These non-trend data points are confirmed when looking at a geo-statistical plot of the same data, found in Figure 5.9.

Figure 5.9 On-Campus New England Geostatistical Analysis (10 ms Acceptability Range)

5.3 Statistical Psychotherapy (Breakdown)

The two previous sections gives insight into the arrangement of the data, yet does not go into specifics. It is useful to be able to look at the mathematical aspects of the data, and compare each of our datasets to each other in a statistical manner. By performing a series of statistical analyses, we were better able to understand what our data was truly telling us. These analyses gave us a much clearer picture as to how accurate our data actually was, and how meaningful the conclusions we drew from this data were. In performing these different manipulations of the data, we were able to see exactly what our data was telling us.

On-Campus Nationwide			*Off-Campus Nationwide*	
Mean	48.4153		Mean	67.87432
Standard Error	2.003528		Standard Error	2.116667
Median	48		Median	63
Mode	50		Mode	63
Standard Deviation	27.10323		Standard Deviation	28.63373
Sample Variance	734.5848		Sample Variance	819.8907
Kurtosis	-0.57239		Kurtosis	-0.22653
Skewness	0.300975		Skewness	0.604027
Confidence Level(95.0%)	3.953129		Confidence Level(95.0%)	4.176361

Table 5.1 On and Off-Campus Nationwide Statistical Analysis

To begin, we analyzed the Nationwide datasets, for both the on- and off-campus data. As can be seen in Table 5.1, the statistics for this pair of datasets do not vary greatly. The mean, median and mode are within 15 units of one another. The standard deviation and standard error vary much the same way, with a gap of only 1 and over 100 respectively. We found that the data matched what we expected, with the standard deviation, mean and median values for the on-campus dataset lower than that of the off-campus set. We believe that this is possibly a result of the WPI's Internet2 connection, or another connection which the school uses. This increased speed would keep the on-campus data more uniform, reducing the standard deviation, and would keep all the overall values lower than that of the off-campus dataset.

Other important statistical traits which are useful in analyzing our data are the skewness and kurtosis values. The skewness value measures the symmetry of the data collected. The kurtosis value measures weather the data is peaked or flat, relative to a normal distribution. A high value in this measurement means that the data has a very high peak, with rapidly decreasing tails on either side. A low value means the opposite, a flat top with slowly decreasing tails. The best method for analyzing these values is a histogram, and so we created individual histograms for each dataset, which appear in

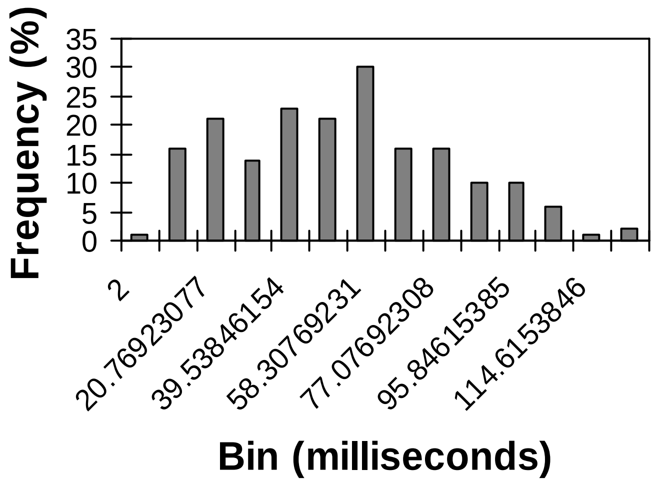

Figure 5.10 and Figure 5.11. The x-axis represents the bins, which are automatically generated in Excel, and represent statistical divisions of the data. The frequency, represented on the y-axis, is the percentage of data points which fall into each bin.

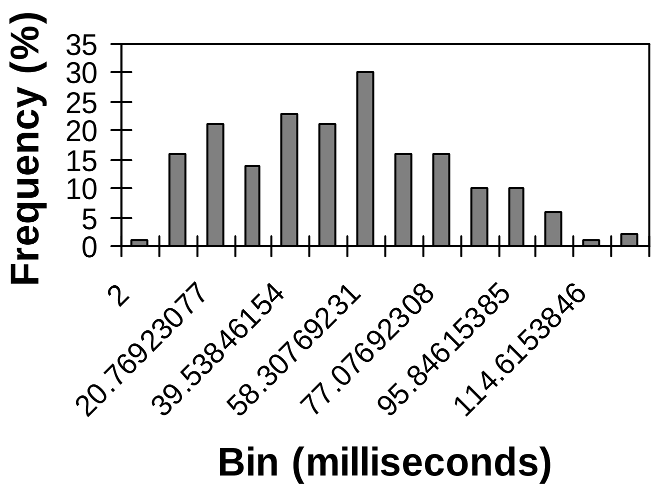

Figure 5.10 On-Campus Nationwide Histogram

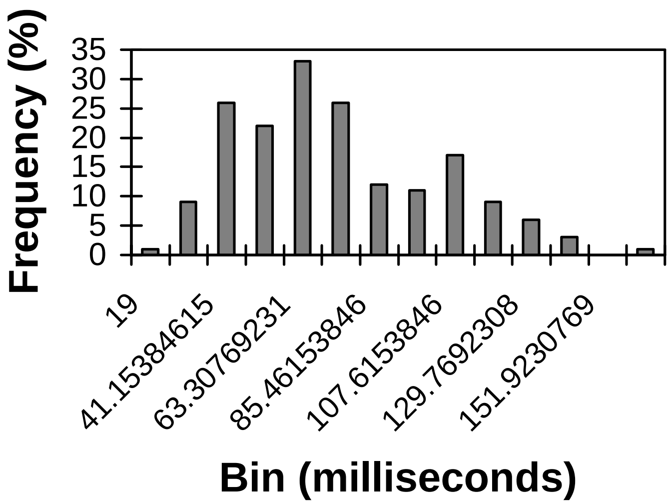

Figure 5.11 Off-Campus Nationwide Histogram

Using the two histograms above, it can clearly be seen that for the nationwide data, there is a fit much like a normal distribution. This fit is only slightly different than the bell curve we were expecting, and tells us that our data is somewhat accurate. Most of the data seems to be shifted towards the low end of the graphs, with only a few outliers on the far ends of the spectrum. This simply shows us that, though there are a few outliers, the data does follow a trend, and is most likely a linear grouping.

After completing this analysis for the nationwide datasets, we then performed the same analysis to the smaller New England datasets as displayed in in Table 5.2.

On-Campus New England	
Mean	22.02381
Standard Error	2.796316
Median	17
Mode	5
Standard Deviation	18.1222
Sample Variance	328.4141
Kurtosis	0.994109
Skewness	1.315351
Confidence Level(95.0%)	5.647274

Off-Campus New England	
Mean	36.57143
Standard Error	2.651264
Median	31
Mode	31
Standard Deviation	17.18216
Sample Variance	295.2265
Kurtosis	0.010393
Skewness	0.696144
Confidence Level(95.0%)	5.354337

Table 5.2 On and Off-Campus New England Statistical Analysis

These datasets were similar to the previous nationwide sets, but for a smaller sampling, focused solely on the surrounding New England area. The means for both are within 15 units of each other. The standard error for each varies by less than 1, a very close value to those of the nationwide sets. The respective standard deviations are also close, with a gap of approximately 1 between the two. The variance is somewhat more spread out however, with a distance of 100 or so between the two values.

In terms of Kurtosis and skewness, the datasets vary widely. The On-Campus set has a very low Kurtosis, which is much lower than the expected value of three for a normal standard deviation. The off-campus set also has a very low value for Kurtosis, and so is further from the normal distribution we were expecting. Both of these values indicate that the histogram for these values should be a flat topped image, with slowly sloping sides.   However, this was not the case, as our histograms pictured in Figure 5.12 and Figure 5.13 don't really show any sort of trend, with large gaps in the overall path of the graph.

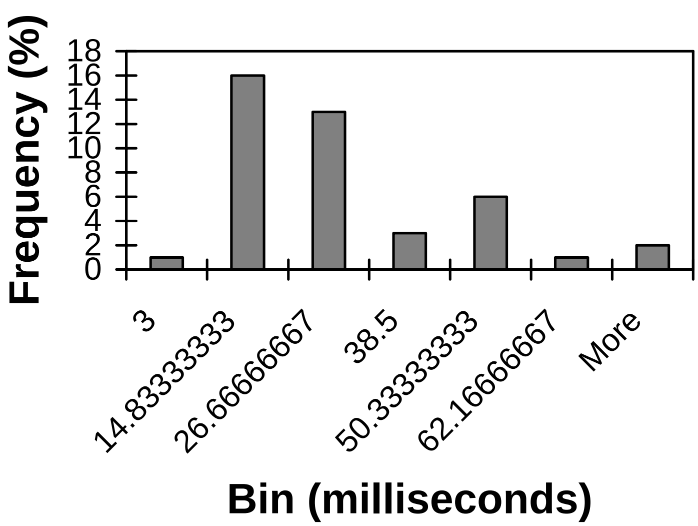

Figure 5.12 On-Campus New England Histogram

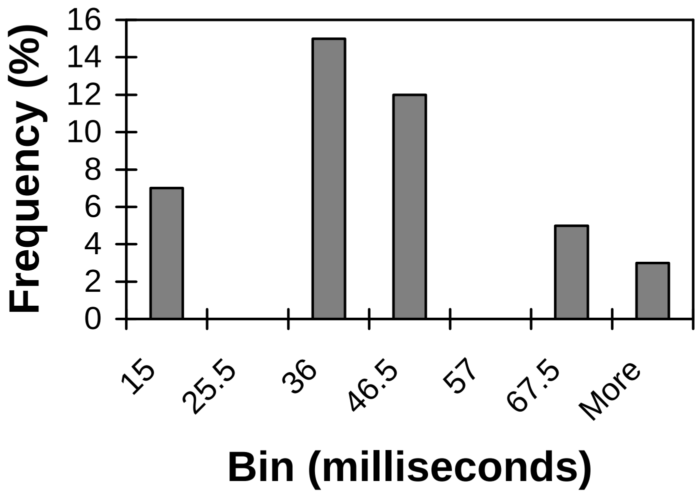

Figure 5.13 Off-Campus New England Histogram

Our datasets seem to show the same trends, with most of the data clustered at the middle-left of the histograms, extending out to the fewer, non-trend data points on the far end. The other statistical analysis we performed shows that our data is close to identical across the individual sets of national and New England data, and also across all of the datasets. This statistical analysis is a good indicator for the accuracy of our datasets and any of the conclusions drawn from these data points.

5.4 Routing Comparisons

There is another element to connectivity that has not been discussed yet. The numbers offer a lot of insight into connectivity and comparing physical connectivity to cyberspace connectivity. Behind these numbers are the paths and patterns of travel. The next analysis has to do with comparing the routes in each case.

Figure 5.14 Routing Comparison for www.nmhu.edu

Figure 5.14 shows the physical travel route (black line) and the digital travel route (gray line). The gray triangle marker is our beginning location: Worcester, Massachusetts. The white square marker is our destination location: Las Vegas, New Mexico. The server found in this city is www.nmhu.edu, New Mexico Highlands University. In our testing, we found that this server had an almost perfect fit to our trend line. The data depicted in Figure 5.14 was gathered from on-campus.

The black line represents the physical travel path used to get from WPI to NMHU. The path depicted is road travel from WPI to Boston MA, air travel from Boston MA to Albuquerque NM (with a connection in Cincinnati OH), and then road travel from Albuquerque NM to NMHU. In order to keep the creation and presentation of the maps simple, we express driving distance with a straight line instead of including the zigs and zags of road travel.

The gray line represents the cyberspace travel path used to get from www.wpi.edu to www.nmhu.edu. There are many more connections on the way to the destination in cyberspace than in physical space. The path we take in cyberspace is from Worcester to Boston, to New York City, to Chicago, to Kansas City, to Denver, and finally to our destination in Las Vegas.

The differences between the two routes are pretty apparent, but what likenesses can we see? One likeness that asserts itself in almost every route comparison we observed for our project is the connection from Worcester to Boston. In both routes, our traveler (person or packet) will find its way into Boston before heading off in the direction of the destination. The reason for this behavior is the fact that Boston is relatively well connected, both physically and digitally. Therefore, traveling into Boston is relatively low cost in relation to the ease-of-travel gained from the strong connectivity of Boston.

Another likeness shared between the two routes is the inherent desire to exploit high speed connections. Just as airplanes travel much faster than cars, Internet backbones transmit information much faster than standard lines. So long as we can maintain a moderately direct route, its common sense to always opt for the faster connection. Because of this, most of the distance covered is done in airplanes and backbones.

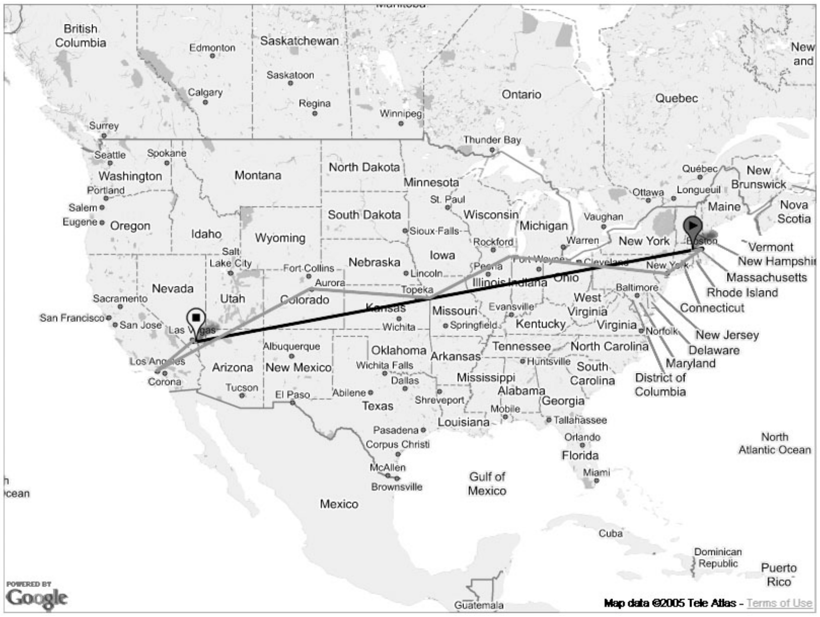

Figure 5.15 Routing Comparison for www.nsc.nevada.edu

Now that we have looked into routes for a point that fit well to our trend line, we will take a look at a point that fell above the line. In Figure 5.15, we see www.nsc.nevada.edu. This server had a relatively high ping in relation to its physical travel time. The physical path is as simple as it could be. A traveler drives to Boston MA, and then flies directly from Boston MA to Las Vegas NV. The fact that there is a commercial airport directly in the destination city means that a great majority of our travel time is spent in a high speed corridor. The same is not true for the Internet travel route. The cyberspace route doesn’t have such an easy time connecting to NSC. It seems that the closest Internet backbone hub is in Los Angeles, and we are forced to utilize slower lines to get into Las Vegas. Combining the good luck of the airport and the bad luck of routing through Los Angeles, the result is a non-trend point falling above the trend line.

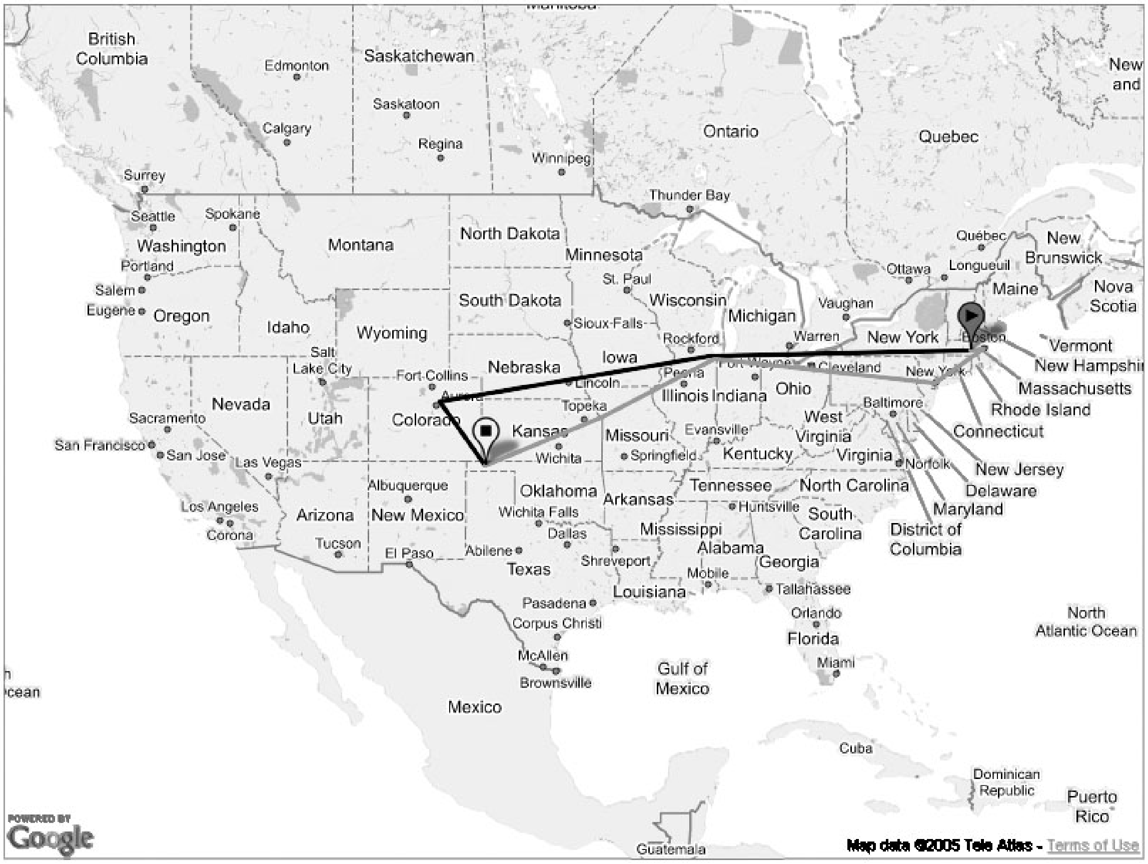

Figure 5.16 Routing Comparison for www.opsu.edu

Our luck with routing is reversed in Figure 5.16. This map is for www.opsu.edu, a server that fell well below the trend line. The Internet trace flows through Kansas City again, just like we saw in Figure 5.15. Unlike in other maps, the Internet data is able to remain on a high speed line all the way to the destination. The data manages to shoot across Kansas in less than 10ms. As a result, the cyberspace connectivity is superb.

The fact that OPSU is a non-trend point is only half dependant on the cyberspace connectivity though. The physical route illustrates very poor physical connectivity. The closest commercial airport to OPSU seems to be in Denver CO. The drive from OPSU to Denver is just over 8 hours. When a traveler is forced to spend so much of its journey on a low speed line, such as driving relative to flying, the consequence is notably bad connectivity. This behavior can be seen in both cyberspace travel (Figure 5.15) and physical space travel (Figure 5.16).

5.5 Summary

After analyzing all of our datasets, we were able to create trend lines and linear regressions that would help to explain what was happening in our datasets. While these trends were interesting to us, we were far more interested in the points which did not fall on our trend lines, rather focusing on the major non-trend data points. This allowed us to examine why different areas were connected differently, and compare how each was connected via the Internet and cumulative travel time.

6 Conclusions

6.1 Data Trending

We have discussed the trends that were expressed in the datasets. We've also discussed the points that did not follow the trends. Non-trend points occurred both above the trend line and below the trend line. Statistics only tell so much of the story though. What is the actual, real-world impact of being 'above' or 'below' the line?

When a point in the dataset falls 'above' the trend line, it is because the RTT for that point is relatively higher than what would be expected for the given x-position. In other words, we assert that points 'above' the line are not as well connected in cyberspace than they are in physical space. When a location has good Internet connectivity, it means that data can travel there quickly and efficiently.

In order for a point to fall well above the line, there must be a notable mismatch between Internet connectivity and physical connectivity. The cause of such behavior is in the type of connections used during travel in cyberspace and physical space. For example, a packet can travel on an Internet backbone for a high percentage of the distance to the destination. In this scenario, we would see good Internet connectivity. This behavior can be seen when comparing ping times between Worcester and Boston, and Worcester and Springfield. To reach Boston, packets seem to take a relatively low amount of time (~3ms). Springfield, although being almost the same distance from Worcester, takes packets much longer to reach (~15ms). There is a high speed backbone that connects Worcester directly to Boston, but there does not seem to be any similar connections to Springfield.

Another cause for a location to be placed ‘above’ the line is if physical connectivity is relatively good. The effects of high speed connections are apparent in cyberspace, but the same patterns are mirrored in physical space travel as well. Instead of traveling along backbones like we do in cyberspace, the physical space “high speed corridor” is airplane travel. Airplane travel allows us to travel great physical distances at a high speed, just like backbone pipes allowed us to do in cyberspace. Therefore, if a location has easy access to a large airport, it will be physically well connected. On the other hand, if a location’s closest commercial airport is moderately far away, then we would have to resort to driving for some time to jump on to our ‘high speed’ connection.

The inverse can be said about points below the line. The high speed connections that were mentioned earlier would play the biggest role in how connected a point is on either scale. Points that fall below the line might be very close to or have their own commercial airport. Also, if the location is not near any Internet backbones though, we will find that packets are struggling to travel through the lower speed wires.

With such wild variability in access to high speed connections, why do we see reliable trends in our nationwide datasets? The reason for this is in the nature of the development of connective technologies, such as Internet backbones and air routes. Both physical connectivity and digital connectivity are fueled by the private sector. Because of this influence, supply and demand play a heavy role in the development of connective technologies. There is more demand for a commercial airport in a large city like Denver than there is in a small suburban town. Similar supply and demand applies to Internet development. A company will see more business on a high-speed line added in Chicago where there is a very high population density as opposed to the heart of Montana where the population density is very low. Due to these development patterns, airports and Internet hubs have a tendency to be constructed around the same cities.

6.2 Social Implications of Connectivity

The fact that certain areas are better connected than others is not a novel thing to say; we know this instinctively. Meaning comes when we reason why certain areas are better connected.

6.2.1 Social Causes for Differences in Connectivity

Our early research helped to deliver such meaning. As demonstrated in our results, there seem to be pockets or clusters of nearby servers that have similar round trip times. This can easily be related to our early background research in Section 3.1.1 with regard to the “principle of least effort” [2]. Each stage in the American transportation system grew out of the previous stage. Canals were built to connect rivers and other bodies of water, and then railroads were built in place of filled-in canals. Native Americans made their roads according to the habitat that they lived in, and the early settlers’ roads were built on top of these. Cities grew in place instead of tearing down these roads and building new ones.

The Internet grew in a similar fashion. Pipes were originally laid down to connect important hubs using government wires. As the Internet grew, these lines were integrated with the existing phone system. The telephone lines were already serving residential areas, and so, for small local areas, provided connectivity patterns that are similar to road travel patterns. . The Internet was not expected to grow in size at the rate that it did, and so new connections were made to meet areas of demand instead of being planned in advance in the form of a grid system.

This pattern results in patches of heavy connectivity mixed with patches of light connectivity. Some areas have grown quickly, while naturally other areas have not. These growths often come out of demand, though it is possible for a lack of Internet growth in a certain area to have a negative developmental effect on the surrounding area. It is possible for poorly connected areas to not attract the sort of growth that well-connected areas might attract.

6.2.2 Effects of Connectivity on Business

The development of the Internet in major metropolitan centers greatly affects the development of any sort of commercial or industrial growth. In our increasingly connected age, businesses are constantly looking to remain up-to-date and on top of their competition. If they can have faster, more reliable methods of tracking inventory across multiple cities, or reducing the amount of paperwork that is sent through out their company, they can get the advantage. In this way, cities need to maintain their Internet infrastructure at a high level, to keep pace with growing corporations.

If a city should fall behind in such a technological race, it could soon see its biggest commercial establishments moving to better technological climes. As we found in our research, most of the best connected areas are places which are major cities. This is especially true for our connection to Boston. As explained earlier, every time we ran our tests, our Internet travel paths would always take us through the Boston hub, before heading out to other routers. This exceptionally fast connection only highlights the fact that Boston is a major metropolitan area, which draws business from all across the country. Also, our connection throughout the East coast was exceptionally good, with a corridor of fast RTTs running from Boston down to Washington DC. As most, if not all of these cities are home to numerous major corporations, this high speed connection aids them in intercommunication with their myriad offices.

In the same way, if physical connections around a city like Boston were much poorer, fewer large businesses would flock to that location, reducing the overall growth and expansion of the city as a whole. If people are unable to get into the city to work, or live in the city and are unable to get out, this causes a major problem, which will also drive away businesses which could provide much needed boosts to burgeoning populations. If people cannot get to the businesses, they cannot purchase anything, and so eventually drive the store or shop bankrupt.

6.2.3 Effects of Connectivity on Everyday People

For a normal person, the effects of connectivity for both Internet and physical routes can be large. For many, the use of the Internet and the tasks which it helps to perform are extremely important. Many college students use it for research, or for entertainment. If the connectivity of the Internet for an individual is affected in any way, it can drastically affect the productivity and often times the livelihood of the person involved.

In the past, slow connections to the Internet such as dial-up were the only methods available to access the information on the World Wide Web. People accepted this as the norm and were able to increase their productivity. As connections evolved, they grew faster and faster, increasing the amount of work that could be done through the

Internet. Many offices allow their employees to work from home, and these employees are able to access the necessary information they require through just their computer.

This, however, also leads to problems, as slow connectivity can drastically affect any kind of user. As can be seen through our data, people working and living on the East coast seem to have better connections, or at least are closer to the high-speed backbones. As our examination proceeds westward, the time it takes to reach specific points slows down greatly. The most probably reason for this slowdown is the fact that many rural places, such as North Dakota, have much poorer networks than more heavily settled areas. This means that for someone attempting to connect from back-woods Montana to a server in Massachusetts the times will be much higher, as they are further from the high-speed backbones that connect places like Boston, New York, Chicago and Los Angeles together. The fact that the network is less developed is a result of the lack of personal demand for better connections.

6.2.4 Pockets of Connectivity and Their Effect on Society

As can be seen in Figure 5.5, there exist several bands of connectivity across the country that spread out from New England in what almost resemble waves. From this figure, one can make general assumptions about Internet pocket location and size. There seems to be a relatively well-connected pocket of locations in the Chicago area, as well as the Salt Lake City area. The same map shown in Figure 5.7 with geostatistical analysis gives a slightly different impression. Here you can see pockets of strong connectivity and weak connectivity. An example of a strongly connected pocket would be the locations around the Baltimore area; these locations would be part of a strong Internet corridor that goes from Massachusetts to Washington, D.C. The locations directly to the west of these locations in the Kentucky/Tennessee area appear to be poorly connected. This is possibly due to a strong backbone connection that stretches from Boston to Chicago but bypasses points in between. This can result in locations that are passed over for Internet connectivity. Locations that are relatively poorly connected also can be found in Vermont, which is relatively close to Massachusetts. This would imply that although Vermont may be well connected physically, it is not as well connected with regard to the Internet.

We conclude from these results that not only is there a correlation between physical connectivity and Internet connectivity, but that locations with similar connectivity are often located close to one another. The existence of pockets and clusters of connectivity implies that entire regions can be well connected or poorly connected, not simply individual locations.

Society affects connectivity in a sort of circular feedback loop. As a particular region grows, its demand for connectivity grows. As connectivity grows, it enables the region to grow further, and the process repeats. It is possible that clusters or pockets of connectivity exist to satisfy the needs of those living in that region. The connectivity needs of those in large cities are conceivably greater than the connectivity needs of those in areas that are more rural. Because of this higher demand in cities, connectivity spreads more quickly in areas of high population density than in areas of low population density.

The fact that good or bad connectivity is associated with entire regions rather than individual locations has a profound effect on society. First, it further emphasizes the feedback loop concept discussed in the previous paragraph. Second, it allows those seeking a certain level of connectivity to find it within a broad range of points in a particular geographical area. Otherwise, if points within a particular region did not share similar connectivity traits, it would be difficult for those who desired a certain level of connectivity to choose such a location. It would also make it more difficult to measure connectivity: one would not be able to make inferences regarding the connectivity of any point even if it were close to a point with known connectivity.

6.3 Connectivity Comparisons

Through our work and analysis of the data collected, we were able to create several conclusions about our methodology and comparisons. First, after comparing the RTT to Cumulative Travel Time and to driving distance, we found that the correlation between the latter was far higher than the CTT. This puzzled us, as we believed that the CTT, more accurately reflecting a similar routing as RTT, would have a higher correlation. Alas, this was not the case, with both the comparisons of RTT to Driving time and driving distance were better indicators of correlation.

We find a reasoning which would allow us to understand why this is happening. Perhaps it is due to routing differences in both the Internet times and the physical space times, or the fact that we have to run the data through numerous scripts that could be causing this large difference. This difference could also be caused by incorrect data being passed into or analyzed by our scripts, which would throw our data off immensely.

Also, in comparing our nationwide datasets to our New England sets, we found other interesting points which drew questions. Our New England data showed almost no trend at all, which we at first thought could be a result of our focus being so tight to such a small area. We also believed that the lack of correlation could have been caused by other, more far reaching factors such as differences in the number of data points we were using for each test. The New England dataset was a smaller number than the nationwide, and included many more points focused solely the New England area. Overall, we are unsure as to why our correlations vary so greatly, and believe that further investigation of this in a more comprehensive manner than we were able to complete is needed. For more information about further work, please refer to Chapter 7.

6.4 Effects of Routing

When we were discussing routing, we saw that proximity to backbone hubs and airports played a huge role in how connected a location is. But what conclusions can we draw from the routes themselves?

We saw some similar routing behavior in both the cyberspace routing and the physical space routing. We also saw that behavior shared across the three maps (Figure 5.14, Figure 5.15 and Figure 5.16). For this reason, we assert that routing can be used a metric to discuss connectivity between two locations. In the case of Worcester and Boston, we saw that Boston offered Worcester lower cost connections to the destination than we would have otherwise seen. Similar cases are shown in our data in both physical connections and digital connections.

An important characteristic of Internet routing is that the number of nodes used doesn't seem to play any direct role in how connected a location is. Meanwhile, if there are no direct flights from one airport to another, the connecting of two flights will result in a more difficult and troublesome journey. The reason for such a discrepancy is in the turn around time of the traveler in question. In the case where our traveler is a packet on the Internet, a node can process the packet and shoot it down to the next node on the path. The only waiting that's done is in the queuing before the packet is processed (which is not considered in our data). In the case of connecting flights, an airway traveler can not jump off of one plane and onto another like the packet does on the Internet. Our airway traveler needs to be processed, just like the packet. In the case of the packet, processing is a nearly instantaneous event. In the case of our airway traveler, it is not. Even so, such processing is simply queuing, which was not considered in our data. The real trouble in connecting flights is due to the fact that the next flight simply will not get underway the moment the first flight landed.

Routing is an important aspect of connectivity because true connectivity is a result of considering weights and costs of a travel path. We can only talk about connectivity on an absolute scale when we optimize our route using the highest speed connections and avoiding low speed connections for as much of our journey as possible.

6.5 Summary

In all, our data showed us several clear things. First, the trends of the data showed how mismatches in RTT and CTT could affect the connectivity of a point. Second, we examined the social implications of this connectivity, both on businesses and on society as a whole. While we were able to complete a large amount of analysis, and reach many different and interesting conclusions, there were areas that we were unable to touch upon. These areas of interest offer opportunities for further research and experimentation in this field of study

7 Further Work

7.1 International Datasets

After working with our USA based datasets, we began to look elsewhere for further avenues of exploration. We wondered if there would be a difference between what we found here in the US compared to our connectivity to other countries around the world. By developing a set of university websites in other countries, we could run our scripts and other analysis tools and easily see if there is any difference between our initial research and this new information. We started to compile a list of colleges in other countries, before we refocused our project onto just the US data.

While we chose not to investigate this avenue, another project could easily add to our work by performing such an international analysis. By gathering a list of international college websites, and a list of international airports in each of the respective countries, an analysis much like we completed could be done. Other data that would need to be collected is the latitude and longitude of each of the websites, the latitude and longitude of each of the airports, and each point would probably need to be checked to make sure that they are actually where they say they are. Another group might have to change the online resources we used to get the travel distances, as we were unsure as to whether MapQuest would work for overseas locations. The maps and charts we created using Google Maps would be much the same to reproduce on a larger scale for a world wide dataset.

Potential areas of focus for such a research endeavor could reach in several different directions. One big area of focus would be our connectivity to less developed countries versus countries like England, France, Germany, and other western countries. Other investigations of interest would be our connectivity to the Far East, where the fastest connections can be found, and what routes information takes across multiple countries. All of these ideas would make for many different interesting and exciting projects, which further groups could investigate in-depth.

7.2 Non-Educational Datasets

After analyzing the data we collected for the many different colleges, we wondered what other directions we could take our investigations. Looking beyond just adding more schools to our list, we wondered what would happen if we branched out to look at other websites, like those of businesses, across the country. We looked briefly into this topic, but were unable to find a source which would give us a list of websites where we would actually know the exact location. Without that data, we would not have been able to perform any comparisons similar to what we performed for our project.

Should another group be able to either find or compile a list of non-collegiate websites and find the exact position of the servers that those websites are on, they would be able to perform a different kind of comprehensive testing. They would be able to see what our connectivity is compared to other, non-college sites, and would show how well connected the businesses are to the backbones. This would be an interesting area to compare against our college data, and the analysis of both sets together would provide interesting insights into different types of connectivity.

7.3 Assessing the Impact of Traffic

Data throughput is by no means consistent. There are many variables that can affect the round trip time of a particular request. If the receiving server is fairly busy, it might not be able to respond to a request quickly. Localized Internet traffic can affect routing choices; a packet will always attempt to choose the fastest path, but this path is very complex and may not always be the most direct in terms of distance. General congestion on the Internet will also slow things down, as even backbone pipelines have a maximum throughput. Depending on the time of day, this general congestion can affect overall Internet connectivity to varying degrees. There are predictable patterns of Internet congestion based on history: it is perfectly reasonable that the Internet will be busier at 2:00 in the afternoon than 2:00 in the morning.

Many of these variables have their real-world counterparts. Highways can become fairly congested and one might choose a slower, less optimal path on a backroad that is faster on a given occasion due to highway congestion. If an airline flight is completely booked, or is experiencing severe delays, one might have to book a different flight that takes longer or has a stopover at another airport. All of these factors make it difficult to make consistently reliable judgments of connectivity.

As such, we attempted to minimize the effect of Internet traffic during data collection by pinging all servers three times and keeping track of the minimum ping time. We also pinged all servers several times over the course of a few weeks to minimize the effects of any momentary Internet traffic. We always pinged the servers at night during low Internet traffic periods. When choosing flights, we always looked for flights leaving at least a week in advance in order to increase the potential of finding a fast flight time.

We used MapQuest to derive our driving time estimates, and these were not affected by traffic at all.

7.3.1 Sample Size and Statistical Accuracy

Though we hope that we have minimized the effects of congestion as much as possible, it should be understood that these effects can not be eliminated. Based on the time limitations of our research, however, it is impossible to know for sure the degree that congestion affects our results. A reasonable hypothesis may be that there is a positive correlation between the number of pings performed on a single server and the statistical accuracy of the result. If a server is pinged just once, it is impossible to tell if the response time is reasonable for the server, or to what degree the response time is affected by traffic. However, we have no way of knowing how many times we must ping a server before we have a statistically viable sample. We wonder if more research could be done to prove mathematically that we have conducted enough tests.

7.3.2 Absolute Minimum Response Time

If Internet traffic were truly not a factor, response time could then be measured with pinpoint accuracy. In this case, Internet travel time could be measured in a similar manner as with driving time on MapQuest, based on the fastest direct route. If one was to reduce the Internet to a weighted node graph and calculate for each passage on the Internet the associated distance and maximum throughput, one could calculate the absolute minimum response time physically possible. This data could be then compared to the results we achieved in real world testing.

7.3.3 Real World Traffic Congestion

Though we measured our Internet travel time by experiment, our physical travel time data was collected by using given MapQuest and Expedia.com travel time algorithms. Though we have discussed minimizing the effect of traffic on Internet travel time in the two subsections above, it might be interesting to use real world data for our physical travel time and therefore introduce congestion into our physical model as well. Though it might prove difficult to calculate driving time to many different locations, it might be possible to rely on previously collected data. Companies such as taxi, limo, or shuttle services probably make many trips to and from the airport and keep records of travel time as well. It is also probable that airports would collect actual travel times. While such a project may be possible, we are unsure how feasible such a project might be, or how interesting the results might be.

7.4 Alternative Data Comparisons

Our project focused on the round trip time to a point compared to the actual travel time to that location. There are other metrics, besides RTT, that could be used to measure the Internet connectivity that we did not use, but could offer different results than what we found. These include any of the metrics suggested in our background section, which are IP path length, Autonomous System path length, or simply straight-line versus routing distance to the server. Any of these would provide a different result than just pure RTT, and could potentially ask or answer many new questions that didn't even occur to us as we were working on our methodology.

Also, as we proceeded with our investigations, we noticed, as discussed above, that other comparisons seem to have a better correlation than our chosen method of comparing RTT to Cumulative Travel time. When we compared the RTT vs Distance, and the RTT vs Driving Travel Time, we found the correlation matching better than our method. We did not have enough time to investigate this phenomenon, but a further study of this strange data would be an interesting assignment for another project. If further rigorous testing could be completed, perhaps the answer as to why the correlation is so much greater could be found.

A whole project could be devoted to just finding the metric that has the best correlation, but we did not have the time to develop this during our project. Should this topic be investigated, our data would be a good base from which to start from, as we already began some of the comparisons and briefly gave a superficial look at the correlations of several different metric comparisons.

7.5 Correlation Investigation

In all of our comparisons between Internet travel time and real-world travel time, we chose to compare round trip time to cumulative travel time. When people travel to a physical destination that is far away, they will (issue of cost aside) drive to an airport, fly to the airport nearest their destination, and then drive the rest of the way. The Internet is largely similar; due to the existence of high-speed backbones that allow data to pass through certain corridors rather quickly, it makes the most “sense” for packets on the Internet to find the quickest path to such backbones in order to get to their destination as fast as possible.

For each location in our dataset, we measured or otherwise kept track of the driving time directly from WPI to the location (as reported by MapQuest). In doing so, we also determined the distance from WPI to the location. In early statistical analysis, we plotted the driving time and distance against the round trip time for each location, and then constructed linear trend lines. This produced three trend lines in total, comparing round trip time to driving time, distance, and total cumulative travel time. It was the latter of the three that we decided best modeled our real-world understanding, and as such was the relation that we built our conclusions on. However, after constructing the trend lines for each relation, we found that the cumulative driving time regression provided the lowest r^2 value. The r^2 value is the square of the correlation coefficient; values range from 0 to 1. 0 implies no linear correlation, while 1 implies a perfect linear match.

The implication of this r^2 value is that although we consider total cumulative travel time to be our best real world model, it provides for the worst mathematical predictor of Internet travel time. It could potentially be the case that there is a much higher gain in speed when transitioning from a car to a plane than there is when transitioning from a local Internet route to a backbone. Perhaps the speed gain from an interstate is a more realistic model for the speed gain of a backbone. Perhaps there is another cause that we aren't aware of. Because we discovered this issue when analyzing our data, we were unable to dedicate considerable time to this question. We believe much more research could be conducted in this area, and are happy that our conclusions can generate new questions in addition to answers.

7.6 Routing

Although we have looked into some aspects of routing, our previous explanations were somewhat superficial. Routing is a tricky entity, especially in cyberspace. A packet can travel one path at 5:00am, and then travel a very different one five minutes later. We also only looked at two types of routing and simplified it for ease of visual comparison and ease of creation.

There are many unanswered questions and unexplored paths in routing. If roadways were included in our routing maps, we might have been able to compare back roads to high ways. We also could have explored traveling with railways. In some cases, we might have seen a long drive between an airport and a location. But there may have been a direct railway connecting the city with the airport and the destination city.

An important part of routing in both the physical world and cyberspace is queuing. In Internet routing, it is possible for a node to queue a packet for longer than it took the packet to travel there. We can see similar behavior in airports, but not to such an extreme. Since a study including queuing would have taken some very difficult and most likely subjective testing, we decided to not take it in to consideration.

We mentioned that Internet routing can actually change during different times of day. Such behavior is a result of traffic. If a project wanted to look at changes in routes though, it would be possible to do so without having to factor in traffic. Changes in routes offer a lot of interesting discussion even though it is only a side effect of traffic.

7.7 Summary

In conclusion, this type of research can take many paths. While we chose to focus our efforts on the comparison between RTT and CTT, there are other ideas that could be investigated further. By looking into international or non-educational datasets, the same type of experiment could be conducted to compare the RTT and CTT for each set. Other areas of interest could include investigating the impact of Traffic on the data that is collected, which could spawn a whole list of other topics that would be related. Also, by utilizing other data comparison metrics, like RTT to distance or others, could provide a different picture of what connectivity really is across the country. Finally, while we touch upon routing in our conclusions, much more research could be conducted to examine this area in depth

8 Bibliography

1. AnalogX, Internet Traffic Report, http://www.internettrafficreport.com
2. Black, William, *Transportation: A Geographical Analysis*, The Guilford Press, New York, NY, 2003
3. Cairncross, Frances, *Death of Distance*, Orion Publishing Group Limited, London, 1997
4. Claffy, K. Skitter, http://www.caida.org/tools/measurement/skitter/ .
5. Computer History Museum, http://www.computerhistory.org/timeline/
6. Dodge, Martin and Rob Kitchin, *Atlas of Cyberspace*, Pearson Education, 2002.
7. Flournoy, Don, The Last Mile: Where Telecommunications Slows to a Crawl, Ohio University School of Telecommunications, Athens, OH, http://www.tcomschool.ohiou.edu/its_pgs/lastmi.html#bios
8. Fu, L., An adaptive routing algorithm for in-vehicle route guidance systems with real-time information, (Department of Civil Engineering, University of Waterloo) Source: Transportation Research Part B: Methodological, v 35, n 8, September, 2001, p 749-765.
9. G'unther, Andr'e, Christian Hoene, Measuring Round Trip Times to Determine the Distance between WLAN Nodes, Telecommunication Networks Group (TKN), TU-Berlin, Germany
10. http://www.tkn.tu-berlin.de/publications/papers/hoene_paper2.pdf
11. Huffaker, Bradley, David Moore Marina Formenkov, Daniel J. Plummer and K. Claffy, Distance Metrics in the Internet, http://www.caida.org/outreach/papers/2002/Distance/distance.pdf , 2002.

12. Lepak, Joel, M. Crescimanno, Speed of light measurement using ping, Center for Photon Induced Processes, Department of Physics and Astronomy, Youngstown State University, Youngstown, OH, 44555-2001

13. Moore, David, Ram Periakaruppan, Jim Donohoe, Where in the World is netgeo.caida.org?, http://www.caida.org/tools/utilities/netgeo/

14. Moulton and Associates, *The American Transportation Problem, George Banta Publishing Company, Menasha, WI, 1933.*

15. Periakaruppan, Ram and Bradley Huffaker, GeoPlot: About, June 21, http://www.caida.org/tools/visualization/geoplot/ 2005.

16. Periakaruppan, Ram Gtrace – Geographical Traceroute, http://www.caida.org/projects/internetatlas/ .

17. Wikipedia.org, Routing, http://en.wikipedia.org/wiki/Routing. 2005

18. Wikipedia.org, Transport, http://en.wikipedia.org/wiki/Transport. 2005

Appendices

Below are the data described in Section 4.1 for both the national and New England sets. In order to conserve space so that all fields could be included in the width of a single page, the units of measurement were not included. Those units are as follows: CTT (minutes), driving time (minutes), distance (miles), latitude (degrees), longitude (degrees), and RTT (milliseconds).

Appendix A National Dataset

College	Zip Code	Airport Code	CTT	Driving Time	Distance	Latitude	Longitude	On-Campus RTTs			Minimum On-Campus RTT	Off-Campus RTTs			Minimum Off-Campus RTT
alamo.nmsu.edu	88310	ABQ	687	2122	2262.47	32.7929	-106.161	67	67	67	67	107	142	114	107
lasvegas-college.com	89044	LAS	449	2434	2701.71	35.9577	-115.1586	50	51	51	50	68	87	85	68
oregonstate.edu	97330	PDX	640	2828	3138.9	44.6364	-123.2804	91	88	88	88	187	122	127	122
weber.edu	84201	SLC	431	2089	2327.66	41.2444	-111.9977	62	63	62	62	112	148	134	112
www.aib.edu	50301	DSM	356	1205	1265.66	41.6049	-93.6319	72	73	72	72	97	113	101	97
www.aju.edu	35201	BHM	329	1104	1151.18	33.5191	-86.8093	57	57	58	57	63	81	63	63
www.albanytech.org	31701	ATL	416	1202	1237.78	31.5622	-84.1659	131	46	41	41	56	61	58	56
www.americanglobalu.edu	82001	DEN	436	1739	1895.34	41.1301	-104.8689	92	92	96	92	93	109	93	93
www.arbaptcol.edu	72201	MEM	397	1342	1420.44	34.7467	-92.2799	78	76	118	76	92	118	111	92
www.arizona.edu	85701	TUS	527	2415	2702.57	32.2167	-110.9709	69	69	69	69	99	130	117	99
www.augsburg.edu	55401	MSP	272	1303	1356.22	44.9847	-93.2709	53	52	54	52	51	70	60	51
www.barry.edu	33138	MIA	281	1389	1471.62	25.8552	-80.1846	45	45	46	45	58	73	80	58
www.bates.ctc.edu	98401	BLI	753	2758	3089.22	47.2538	-122.4431	81	94	81	81	96	99	84	84
www.baylor.edu	76701	DFW	429	1711	1833.02	31.552	-97.1385	65	60	57	57	64	94	78	64
www.bbc.edu	18411	MDT	273	273	264.14	41.4656	-75.7341	9	9	9	9	26	50	42	26
www.bc3.edu	16001	PIT	224	546	535.62	40.9096	-79.9361	34	43	33	33	48	81	72	48
www.bemidjistate.edu	56601	FAR	579	1565	1487	47.5271	-94.7611	63	62	63	62	61	71	62	61
www.bic.edu	21201	BWI	166	390	371.81	39.294	-76.6226	21	16	17	16	30	48	57	30
www.bluefieldstate.edu	24701	CRW	396	746	747.75	37.2994	-81.226	35	33	32	32	45	63	46	45
www.bmc.edu	38610	MEM	342	1290	1281.17	34.6645	-89.001	83	84	83	83	103	118	109	103
www.bmcc.org	49715	FNT	540	942	988.74	46.4104	-84.5584	61	61	58	58	74	87	65	65
www.boisestate.edu	83701	BOI	513	2339	2628.91	43.6231	-116.3203	91	91	92	91	100	123	120	100
www.brandeis.edu	02451	BOS	49	49	44.62	42.3987	-71.2592	4	2	2	2	41	52	65	41
www.brookstone.edu	28201	CLT	205	821	833.29	35.2287	-80.8458	30	27	26	26	41	40	15001	40
www.bts.edu	04401	BGR	114	263	267.7	44.8521	-68.8311	34	28	28	28	62	79	82	62
www.burlcol.edu	05401	BTV	131	241	242.32	44.4902	-73.2253	21	21	49	21	40	69	59	40

www.butler.edu	46201	IND	244	870	912.72	39.7745	-86.1096	31	29	34	29	47	71	63	47
www.byu.edu	84601	SLC	439	2125	2353.83	40.2247	-111.6938	83	66	67	66	97	467	100	97
www.byui.edu	83440	SLC	605	2264	2529.83	43.7568	-111.6247	99	99	98	98	109	129	115	109
www.cazenovia.edu	13035	ROC	268	268	269.58	42.9403	-75.8232	24	31	24	24	39	55	53	39
www.ccri.edu	02818	BOS	67	67	55.9	41.6429	-71.4939	4	5	4	4	19	39	42	19
www.ccv.edu	05201	ALB	153	153	129.23	42.9148	-73.1152	65	67	66	65	71	85	79	71
www.centralgatech.edu	31201	ATL	311	1101	1151.98	32.8993	-83.468	38	38	36	36	49	81	69	49
www.centralstate.edu	45384	CMH	258	766	799.4	39.7163	-83.8778	45	45	45	45	63	84	72	63
www.centuryuniversity.edu	87101	ABQ	466	1998	2198.44	35.1995	-106.6442	79	79	81	79	83	102	84	83
www.chattanoogastate.edu	37401	HSV	467	976	1006.22	35.0459	-85.3097	32	35	34	32	47	63	82	47
www.cim.edu	44101	CLE	194	579	605.37	41.4995	-81.6959	43	180	43	43	40	74	61	40
www.clackamas.edu	97045	PDX	558	2742	3063.41	45.3306	-122.529	86	87	90	86	106	112	221	106
www.clarkson.edu	13676	BTV	307	373	348.38	44.6469	-74.9157	19	20	19	19	50	37	38	37
www.clarksoncollege.edu	68046	LNK	449	1346	1411.86	41.1138	-96.0413	55	54	55	54	75	114	90	75
www.clemson.edu	29631	GSP	360	955	965.62	34.674	-82.8212	43	44	43	43	58	95	78	58
www.cmcc.edu	04210	PWM	172	172	168.28	44.0965	-70.2571	48	45	50	45	64	75	90	64
WWW.conncoll.edu	06320	BDL	82	82	75.74	41.3514	-72.1063	7	8	7	7	59	26	53	26
www.crowder.edu	64850	XNA	440	1366	1452.23	36.8753	-94.3946	50	52	50	50	164	80	63	63
www.cscc.edu	43085	CMH	206	698	732.45	40.0994	-83.0166	43	45	44	43	54	56	68	54
www.cvcc.edu	28601	CLT	265	818	822.6	35.7628	-81.3219	40	39	40	39	53	71	61	53
www.dacc.edu	61832	IND	252	961	1002.5	40.1365	-87.6307	69	35	36	35	88	76	93	76
www.dartmouth.edu	03755	MHT	155	155	152.05	43.7256	-72.2368	15	15	15	15	46	62	46	46
www.delmar.edu	78401	CRP	419	1954	2038.97	27.798	-97.4011	55	55	55	55	62	73	77	62
www.denverseminary.edu	80110	DEN	372	1782	1944.41	39.6467	-105.0092	61	60	61	60	93	109	93	93
www.desu.edu	19901	ILG	320	366	343.85	39.1624	-75.5199	19	18	16	16	34	120	46	34
www.drury.edu	65801	XNA	512	1296	1372.37	37.1312	-93.2917	50	47	46	46	61	80	86	61
www.dtcc.edu	19801	ILG	283	320	302	39.722	-75.5386	18	19	18	18	31	31	46	31
www.dwu.edu	57301	FSD	431	1488	1585.78	43.7034	-98.0626	68	71	62	62	70	84	89	70
www.eac.edu	85552	TUS	641	2384	2645.54	32.7621	-109.8381	79	79	78	78	163	191	224	163
www.ecu.edu	27833	ILM	432	705	691.6	35.6116	-77.3732	23	22	22	22	43	61	59	43
www.elizabethtowncc.com	42701	SDF	339	950	977.68	37.7021	-85.8419	95	97	96	95	116	132	148	116
www.fairfield.edu	06824	JFK	137	137	125.43	41.1692	-73.268	14	12	13	12	36	48	53	36

www.famu.edu	32301	ATL	516	1224	1285.65	30.4294	-84.2579	38	38	37	37	78	62	62	62
www.fit.edu	32901	ORL	311	1239	1302.09	28.0701	-80.6211	38	37	37	37	62	62	62	62
www.fmarion.edu	29501	CLT	343	816	824.41	34.1985	-79.6958	32	31	32	31	43	56	47	43
www.georgetowncollege.edu	40324	CVG	273	876	903.62	38.2421	-84.5534	75	77	74	74	82	105	97	82
www.graceland.edu	50140	DSM	421	1283	1346.18	40.6492	-93.957	71	72	83	71	86	93	90	86
www.greenmtn.edu	05741	BTV	207	207	173.36	43.5266	-73.2052	23	21	21	21	31	48	46	31
www.greenville.edu	62246	STL	309	1056	1108.65	38.895	-89.397	43	43	56	43	54	74	68	54
www.hazcc.kctcs.edu	41701	CRW	492	900	888.38	37.2415	-83.2013	44	48	45	44	65	86	65	65
www.hccfl.edu	33601	ORL	336	1279	1325.41	27.9826	-82.3401	49	49	49	49	72	80	61	61
www.hlg.edu	63401	STL	360	1169	1225.02	39.7145	-91.4209	63	63	68	63	73	190	205	73
www.hmc.edu	91711	LAX	495	2618	2917.63	34.1276	-117.7153	90	93	90	90	93	93	93	93
www.ilstu.edu	61761	MLI	467	1008	1051.58	40.5373	-88.9869	33	33	33	33	47	48	53	47
www.indianatech.edu	46801	FWA	293	798	809.59	41.069	-85.1656	40	52	40	40	62	156	46	46
www.itascacc.edu	55730	MSP	477	1472	1416.28	47.2378	-93.5299	68	64	64	64	91	84	94	84
www.iwc.edu	52641	MLI	457	1155	1190.6	40.9915	-91.5765	58	59	77	58	104	101	101	101
www.johnmarshall.edu	30301	ATL	247	1050	1071.75	33.855	-84.3959	44	45	48	44	72	73	81	72
www.jsums.edu	39201	JAN	371	1311	1389.14	32.2896	-90.1841	36	43	36	36	53	69	54	53
www.kilian.edu	57101	FSD	372	1430	1517.14	43.6022	-96.7061	13	19	14	13	57	77	61	57
www.kzoo.edu	49001	AZO	307	790	822.29	42.2663	-85.5617	31	31	31	31	45	69	62	45
www.lakeland.edu	53081	MKE	270	1060	1091.44	43.7072	-87.7387	45	46	46	45	249	144	86	86
www.lincoln.edu	19352	ILG	311	352	322.47	39.7848	-75.891	18	18	20	18	41	58	61	41
www.lptc.bia.edu	68071	FSD	467	1396	1454.1	42.2377	-96.4784	100	106	117	100	107	124	126	107
www.lsue.edu	70535	AEX	487	1555	1645.66	30.524	-92.3908	51	51	63	51	71	88	80	71
www.lsus.edu	71101	AEX	506	1500	1604.46	32.5051	-93.7448	69	69	69	69	136	134	151	134
www.marionmilitary.edu	36756	BHM	427	1201	1227.58	32.6685	-87.3424	55	54	56	54	86	85	59	59
www.martinmethodist.edu	38478	HSV	383	1095	1140.87	35.2173	-87.0259	50	50	50	50	91	79	74	74
www.massasoit.mass.edu	02301	BOS	73	73	65.04	42.0777	-71.0422	12	18	11	11	50	47	55	47
www.mccc.edu	08601	EWR	194	268	244.19	40.2193	-74.7619	19	25	37	19	31	50	53	31
www.mesastate.edu	81501	SLC	678	1997	2178.11	39.0723	-108.5429	66	81	78	66	74	90	88	74
www.milescc.edu	59301	BIL	572	1901	2050.78	46.3777	-105.7778	115	116	116	115	130	143	128	128
www.montana.edu	59715	BIL	574	2124	2370.32	45.6891	-110.9295	97	92	94	92	95	123	110	95
www.moreheadstate.edu	40351	CVG	343	813	835.61	38.2298	-83.4396	46	46	48	46	60	81	102	60

www.mountainstate.edu	25801	CRW	347	709	721.14	37.8169	-81.243	39	41	40	39	52	52	81	52
www.mssu.edu	64801	XNA	463	1361	1443.47	37.1074	-94.5218	50	51	58	50	70	84	80	70
www.msubillings.edu	59101	BIL	458	2008	2229.81	45.6311	-108.3519	99	96	100	96	108	107	129	107
www.mtholyoke.edu	01075	BDL	64	64	55.94	42.2564	-72.5775	8	7	6	6	31	37	38	31
www.mwsu.edu	76301	OKC	505	1655	1793.72	33.9834	-98.4403	54	53	54	53	72	83	85	72
www.mxcc.commnet.edu	06457	BDL	88	88	80.86	41.5474	-72.6585	22	21	19	19	32	54	35	32
www.nashua.nhctc.edu	03060	MHT	61	61	54.26	42.737	-71.4488	26	25	24	24	43	45	65	43
www.nau.edu	86001	PHX	575	2291	2537.19	35.1888	-111.6603	100	100	100	100	100	102	100	100
www.ncat.edu	27395	CLT	291	737	723.06	36.0444	-79.8596	25	26	25	25	46	333	128	46
www.ndscs.nodak.edu	58074	FAR	436	1496	1557.71	46.2825	-96.6087	62	62	62	62	60	79	78	60
www.ndsu.edu	58102	FAR	375	1512	1595.36	46.9259	-96.8507	58	58	58	58	54	96	73	54
www.neiu.edu	60411	ORD	278	897	937.98	41.5054	-87.5907	89	74	38	38	50	47	103	47
www.nemcc.edu	38829	MEM	394	1255	1264.01	34.6741	-88.5234	119	86	82	82	141	118	156	118
www.newmanu.edu	67201	MHK	600	1492	1591.94	37.6897	-97.3414	60	56	56	56	71	94	86	71
www.nic.edu	83814	GEG	542	2427	2736.51	47.6847	-116.7792	99	91	126	91	129	171	143	129
www.nitschools.com	25313	CRW	309	717	732.02	38.4199	-81.7509	51	51	51	51	67	82	83	67
www.njc.edu	80751	DEN	447	1676	1815.52	40.6695	-103.2827	74	71	71	71	105	235	113	105
www.njit.edu	07101	EWR	142	215	191.57	40.7241	-74.1732	14	14	13	13	29	39	73	29
www.nmhu.edu	87701	ABQ	583	1979	2154.92	35.5292	-104.9279	73	77	75	73	134	146	165	134
www.nmmi.edu	88201	ABQ	658	1999	2116.23	33.4293	-104.519	89	80	81	80	119	118	135	118
www.northwestcollege.edu	82435	BIL	577	2079	2233.26	44.8625	-108.9659	86	101	81	81	163	157	120	120
www.nsc.nevada.edu	89009	LAS	456	2438	2704.48	36.057	-114.9608	124	127	135	124	114	131	102	102
www.nsuok.edu	74464	XNA	498	1483	1558.49	35.9078	-95.0047	55	58	56	55	69	103	71	69
www.nunez.edu	70043	MSY	424	1406	1492.03	29.9614	-89.9537	87	90	91	87	97	103	113	97
www.okccc.edu	73101	OKC	399	1534	1655.03	35.4915	-97.5625	48	48	48	48	71	72	60	60
www.olc.edu	57752	RAP	512	1751	1840.89	43.534	-102.3174	68	74	72	68	92	103	110	92
www.opsu.edu	73939	DEN	691	1828	1867.15	36.7696	-101.8105	47	47	46	46	63	71	82	63
www.ovc.edu	26101	CRW	363	691	673.88	39.2786	-81.5101	18	19	22	18	41	63	63	41
www.pacificu.edu	97116	PDX	581	2768	3084.15	45.6684	-123.3405	91	90	92	90	136	149	122	122
www.plymouth.edu	03264	MHT	137	137	135.3	43.8032	-71.7266	14	14	14	14	62	56	71	56
www.potomacstatecollege.edu	26726	PIT	339	532	514.87	39.4294	-79.0096	16	14	14	14	36	39	54	36
www.prescott.edu	86301	PHX	529	2356	2613.23	34.6294	-112.4304	74	74	74	74	108	132	132	108

www.princeton.edu	08540	EWR	185	259	229.05	40.3679	-74.6543	11	10	10	10	38	64	48	38
www.pserie.psu.edu	16501	PIT	294	491	508.41	42.123	-80.0855	33	33	33	33	60	69	89	60
www.ptc.edu	29646	GSP	388	970	973.62	34.1438	-82.1464	44	44	45	44	61	61	71	61
www.rcc.vccs.edu	23149	RIC	230	573	517.39	37.5771	-76.6072	17	17	17	17	65	34	68	34
www.ric.edu	02901	BOS	52	52	41.74	41.8255	-71.4114	6	6	6	6	21	39	37	21
www.richmond.edu	23173	RIC	199	546	522.86	37.5775	-77.5347	23	19	30	19	48	72	41	41
www.roanoke.edu	24153	CRW	467	648	650.19	37.2956	-80.1166	25	24	24	24	65	59	60	59
www.rwu.edu	02809	BOS	76	76	57.39	41.6821	-71.2695	6	6	6	6	21	38	39	21
www.sal.ksu.edu	67401	MHK	523	1476	1571.95	38.8424	-97.6193	53	53	53	53	95	95	103	95
www.salisbury.edu	21801	ILG	394	440	404.87	38.3511	-75.5976	17	17	17	17	35	58	72	35
www.seattleu.edu	98101	BLI	679	2703	3048.17	47.6115	-122.3343	112	109	111	109	111	131	126	111
www.shawnee.edu	45662	CMH	311	797	812.05	38.7891	-82.914	45	47	46	45	55	79	97	55
www.shu.edu	07079	EWR	150	223	196.04	40.749	-74.2606	17	14	14	14	28	46	55	28
www.slcc.edu	84101	SLC	401	2095	2331.74	40.7566	-111.8992	63	62	62	62	117	142	131	117
www.solano.edu	94533	OAK	603	2680	3021.18	38.2845	-122.0168	102	101	102	101	98	116	114	98
www.southeastmn.edu	55987	RST	389	1214	1253.52	43.9823	-91.6349	62	62	62	62	78	272	76	76
www.southeasttech.edu	57101	FSD	372	1430	1517.14	43.6022	-96.7061	50	55	50	50	65	90	15001	65
www.southuniversity.edu	31401	SAV	312	972	998.83	32.0721	-81.0952	52	52	52	52	55	44	54	44
www.stgregorys.edu	74801	OKC	432	1549	1644.28	35.3101	-96.9238	54	53	53	53	203	131	211	131
www.stritch.edu	53201	MKE	217	1007	1038.9	43.0386	-87.9067	41	39	40	39	55	62	66	55
www.sunycgcc.edu	12534	ALB	135	135	129.43	42.2272	-73.7451	16	15	15	15	45	41	46	41
www.sunysb.edu	11790	JFK	194	243	213.44	40.9035	-73.127	10	10	10	10	26	43	46	26
www.suu.edu	84720	LAS	595	2277	2517.78	37.7617	-113.2197	67	67	67	67	103	148	124	103
www.tamut.edu	75501	AEX	592	1470	1566.61	33.3633	-94.2141	60	59	59	59	63	75	92	63
www.tntech.edu	38501	BNA	386	973	1001.02	36.1802	-85.4582	56	56	55	55	79	105	96	79
www.uafortsmith.edu	72901	XNA	477	1480	1574.64	35.3642	-94.4158	75	74	78	74	85	103	102	85
www.uccs.edu	80901	DEN	411	1834	2007.59	38.8336	-104.8206	51	51	52	51	73	92	109	73
www.ucsd.edu	92037	SAN	458	2700	3017.03	32.8548	-117.2497	86	86	85	85	90	91	106	90
www.ucwv.edu	25301	CRW	294	703	719.58	38.351	-81.6265	37	36	37	36	45	63	43	43
www.udel.edu	19702	ILG	286	332	310.44	39.6223	-75.7264	10	10	10	10	40	59	61	40
www.ugf.edu	59401	BIL	715	2265	2323.59	47.5115	-111.2723	73	83	71	71	115	129	122	115
www.uidaho.edu	83843	GEG	604	2555	2854.35	46.7224	-116.9465	94	94	106	94	120	137	120	120

www.uiowa.edu	52240	MLI	404	1107	1155.08	41.6899	-91.4517	39	39	39	39	86	86	86	86
www.umaine.edu	04469	BGR	132	273	277.23	45.0868	-68.6498	15	15	15	15	42	42	43	42
www.umary.edu	58501	RAP	775	1673	1781.26	46.8193	-100.7748	68	68	68	68	81	109	90	81
www.umaryland.edu	21201	BWI	166	390	371.81	39.294	-76.6226	13	14	13	13	31	66	62	31
www.umb.edu	02101	BOS	59	59	51.39	42.3704	-71.0274	7	7	8	7	39	41	39	39
www.und.edu	58201	GFK	380	1574	1666.47	47.881	-97.0607	61	61	61	61	59	80	77	59
www.unh.edu	03824	MHT	110	110	99.74	43.125	-70.9759	10	10	10	10	44	60	61	44
www.unomaha.edu	68046	LNK	449	1346	1411.86	41.1138	-96.0413	49	49	50	49	68	85	89	68
www.uri.edu	02881	BOS	85	85	70.27	41.4796	-71.5207	5	7	6	5	22	38	41	22
www.usi.edu	47701	HUF	212	1022	1065.56	37.9745	-87.5735	37	36	37	36	46	62	46	46
www.usm.edu	39401	JAN	480	1318	1391.67	31.1996	-89.2694	42	41	41	41	63	97	83	63
www.usm.maine.edu	04101	PWM	140	140	138.38	43.6616	-70.2592	11	11	10	10	45	46	235	45
www.utk.edu	37901	GSP	477	872	895.31	35.9646	-83.9197	33	33	33	33	56	66	70	56
www.uwb.edu	98011	BLI	674	2710	3052.88	47.7527	-122.2153	77	77	78	77	87	94	102	87
www.uwplatt.edu	53818	MLI	462	1127	1163.92	42.7466	-90.4931	45	43	42	42	50	58	15001	50
www.uwyo.edu	82051	CPR	680	1821	1973.14	41.5395	-105.8618	54	54	55	54	79	74	80	74
www.vinu.edu	47591	HUF	146	1017	1045.85	38.6298	-87.5042	38	36	36	36	46	46	78	46
www.vtc.vsc.edu	05060	BTV	188	188	186.17	43.9694	-72.6877	74	81	75	74	88	96	128	88
www.washington.edu	98101	BLI	679	2703	3048.17	47.6115	-122.3343	77	77	76	76	86	100	91	86
www.waterbury.uconn.edu	06701	BDL	106	106	98.47	41.5579	-73.0519	6	7	7	6	39	66	44	39
www.wesley.edu	19901	ILG	320	366	343.85	39.1624	-75.5199	28	15	15	15	40	48	47	40
www.wm.edu	23081	RIC	229	593	571.71	37.2081	-76.7746	23	21	24	21	37	71	92	37
www.wmich.edu	49001	AZO	307	790	822.29	42.2663	-85.5617	40	39	39	39	64	101	70	64
www.wncc.edu	89701	RNO	549	2552	2864.44	39.1386	-119.6595	101	101	101	101	99	130	108	99
www.wofford.edu	29301	GSP	321	895	907.86	34.9341	-82.0149	81	82	83	81	93	93	109	93
www.worwic.edu	21801	ILG	394	440	404.87	38.3511	-75.5976	54	120	176	54	59	71	63	59
www.wsc.edu	68787	LNK	568	1452	1481.47	42.194	-97.0263	51	50	55	50	81	98	97	81
www.wyotech.com	82051	CPR	680	1821	1973.14	41.5395	-105.8618	51	52	51	51	67	66	94	66
www3.oakland.edu	48306	FNT	321	676	678.14	42.7216	-83.1464	42	42	41	41	74	82	83	74

Appendix B New England Dataset

College	Zip Code	Airport Code	CTT	Driving Time	Distance	Latitude	Longitude	On-Campus RTTs			Minimum On-Campus RTT	Off-Campus RTTs			Minimum Off-Campus RTT
www.acc.commnet.edu	06082	BDL	71	71	64.36	41.9843	-72.5534	21	22	18	18	46	31	31	31
www.baypath.edu	01106	BDL	69	69	59.49	42.0487	-72.5696	17	19	20	17	31	31	46	31
www.bridgew.edu	02324	BOS	74	74	66.87	41.9726	-70.973	22	5	5	5	62	31	46	31
www.bryant.edu	02828	BOS	52	52	40.3	41.8792	-71.5604	3	3	3	3	15	15	15	15
www.bts.edu	04401	BGR	114	264	267.69	44.8521	-68.8311	29	29	30	29	78	78	62	62
www.burlcol.edu	05401	BTV	131	241	242.32	44.4902	-73.2253	20	21	20	20	31	3093	31	31
www.capecod.mass.edu	02668	BOS	112	112	106.08	41.7107	-70.3611	22	26	18	18	15	15	15	15
www.castleton.edu	05735	BTV	198	198	172.79	43.6401	-73.1505	66	71	66	66	78	93	93	78
www.ccri.edu	02818	BOS	68	68	56.16	41.6429	-71.4939	4	4	4	4	31	15	15	15
www.ccsu.edu	06050	BDL	84	84	76.91	41.6612	-72.7801	15	14	16	14	31	31	46	31
www.ccv.edu	05201	ALB	153	153	129.22	42.9148	-73.1152	60	60	60	60	62	78	62	62
www.clarkson.edu	13676	BTV	307	373	348.37	44.6469	-74.9157	18	18	18	18	31	31	46	31
www.cmcc.edu	04210	PWM	173	173	168.28	44.0965	-70.2571	51	51	47	47	62	62	62	62
www.dartmouth.edu	03755	MHT	156	156	152.04	43.7256	-72.2368	15	15	15	15	46	46	109	46
www.fairfield.edu	06824	JFK	138	138	125.42	41.1692	-73.268	49	134	84	49	46	109	31	31
www.fitnyc.edu	10001	EWR	153	205	180.53	40.75	-73.9967	12	12	12	12	31	31	15	15
www.hartwick.edu	13820	ALB	201	206	211.94	42.4787	-75.0181	52	44	44	44	62	46	62	46
www.hesser.edu	03101	MHT	81	81	72.94	42.9878	-71.4651	45	44	44	44	46	62	62	46
www.jjay.cuny.edu	10001	EWR	153	205	180.53	40.75	-73.9967	17	18	18	17	31	46	46	31
www.keene.edu	03431	MHT	96	96	68.33	42.9764	-72.2744	13	18	16	13	31	31	31	31
www.landmark.edu	05346	MHT	118	118	103.86	43.041	-72.534	20	20	21	20	46	31	46	31
www.mainemaritime.edu	04420	BGR	173	307	292.47	44.3948	-68.7903	39	40	60	39	46	46	46	46
www.massasoit.mass.edu	02301	BOS	74	74	65.04	42.0777	-71.0422	11	23	11	11	46	46	62	46
www.mcintoshcollege.com	03820	MHT	102	102	96.54	43.2027	-70.8909	28	29	29	28	62	3031	31	31
www.mtholyoke.edu	01075	BDL	65	65	55.93	42.2564	-72.5775	5	6	6	5	31	31	31	31
www.mxcc.commnet.edu	06457	BDL	89	89	80.85	41.5474	-72.6585	21	20	19	19	31	31	46	31
www.nashua.nhctc.edu	03060	MHT	62	62	54.26	42.737	-71.4488	23	22	21	21	31	78	46	31
www.nmcc.edu	04769	BGR	286	428	424.97	46.6358	-67.9882	50	54	55	50	62	62	62	62

www.norwich.edu	05663	BTV	179	201	199.04	44.1563	-72.6732	19	20	19	19	78	62	78	62
www.plymouth.edu	03264	MHT	138	138	135.3	43.8032	-71.7266	29	16	14	14	46	62	62	46
www.ric.edu	02901	BOS	53	53	42	41.8255	-71.4114	4	4	4	4	15	15	15	15
www.risd.edu	02901	BOS	53	53	42	41.8255	-71.4114	16	15	15	15	46	46	46	46
www.rockportcollege.edu	04856	BGR	210	246	216.71	44.1792	-69.0943	28	29	29	28	31	31	46	31
www.rwu.edu	02809	BOS	77	77	57.65	41.6821	-71.2695	6	6	6	6	31	15	15	15
www.sunycgcc.edu	12534	ALB	136	136	129.42	42.2272	-73.7451	15	16	15	15	62	31	31	31
www.sunysb.edu	11790	JFK	194	244	213.45	40.9035	-73.127	9	10	10	9	31	15	31	15
www.trincoll.edu	06101	BDL	72	72	66.22	41.7826	-72.6613	5	5	5	5	31	15	15	15
www.umb.edu	02101	BOS	60	60	51.55	42.3704	-71.0274	11	8	8	8	40	44	41	40
www.uri.edu	02881	BOS	86	86	70.53	41.4796	-71.5207	6	5	6	5	15	31	15	15
www.usm.maine.edu	04101	PWM	141	141	138.37	43.6616	-70.2592	11	11	12	11	46	46	46	46
www.vtc.vsc.edu	05060	BTV	189	189	186.17	43.9694	-72.6877	76	74	74	74	78	93	93	78
www.waterbury.uconn.edu	06701	BDL	107	107	98.47	41.5579	-73.0519	6	6	6	6	46	31	46	31